1986

Properties Of Logistic Regression Models With Correlated Observations

John Joseph Koval

PROPERTIES OF LOGISTIC REGRESSION MODELS WITH CORRELATED OBSERVATIONS

by

John J. Koval

Department of Statistical and Actuarial Sciences

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
December 1985

ISBN 0-315-29484-1

ABSTRACT

The correlated logistic regression model, a new model for correlated binary observations in the presence of covariates, is introduced and its relationship with other logistic models established. Comparisons are made with other models for correlated binary outcomes, particularly those that allow unit-specific covariates. A conditional model is derived and shown to be equivalent to a well-known model used in case-control studies.

Properties of several estimators of the regression parameter are investigated. Consistency of the unconditional and conditional maximum likelihood estimators is established and expressions derived for the asymptotic variance of the two estimators under the full model and also under a simpler model. Two other estimators, namely, the estimator obtained by using the usual logistic model and an estimator obtained by using dummy variables in the usual logistic model, are investigated. These two estimators are shown to be, in general, asymptotically biased. Some conditions are given for consistency of the two estimators. Expressions for the asymptotic variances of the two estimators, and the misspecification factors for their estimated variance, are obtained.

In the case of a single binary covariate, it is shown that the bias of the usual estimator is large even for moderate sample size, whereas the bias of the dummy variables estimator becomes small for moderate sample size. It is shown that, although in special cases the usual and

dummy variables estimators may have higher asymptotic relative efficiency than the maximum likelihood estimator, unless both the intraclass correlation of the covariate and the residual intraclass correlation are close to zero, the usual, dummy variables and conditional estimators are, in general, not highly efficient. In particular, the asymptotic relative efficiency of the conditional estimator may be quite low.

Examples are given with simulated and real data showing how the usual, dummy variables and conditional estimators produce falsely significant test statistics.

ACKNOWLEDGEMENTS

I should like to thank Allan Donner for encouraging me to undertake this thesis and for his guidance and supervision, particularly during the early days of this work. Reg Kulperger has been instrumental in obtaining some of the theoretical results on this thesis, and his efforts are greatly appreciated. Also appreciated are the comments of Walter Hauck of Northwestern University, who also provided access to material from his book prior to its publication.

I should like to thank Janet Williams for typing part of the thesis and for guidance on the format of the remainder, and Les Kwarciak for maintaining the computer software and hardware required for the production of this thesis.

Finally I owe much to the confident support of my wife Susan and of our families and friends.

This research was funded in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministry of Colleges and Universities of Ontario through its Graduate Scholarship Program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

General background and review of the literature

1.1 Introduction

This thesis introduces the correlated logistic regression model as a tool for modelling intracluster correlation in the presence of a binary outcome (dependent) variable and covariates (explanatory variables). This model is used to determine the effects of the intracluster correlation on the estimation and testing of the regression coefficients of the covariates. In particular, we examine the properties of maximum likelihood estimators as well as other estimators which have been used in an attempt to adjust for the presence of intracluster correlation without explicitly modelling it.

Section 2 of this chapter gives definitions of terms which occur throughout this thesis. Section 3 covers the origins and development of the logistic bioassay model, the logistic regression model, various multivariate logistic models and finally the correlated logistic regression model. Section 4 considers the use of models to represent the correlation due to clustering and how they relate to correlated logistic models. Section 5 reviews the more complex clustering models which involve covariates and compares these with the multivariate and correlated logistic models. Section 6 examines related models for handling correlation, in particular dummy variables and conditional models, and section 7 considers other work about inference under incorrect models, in particular, in the case of continuous regressor variables in time series, sample survey work, and in case-control studies. Section 8 discusses the regularity conditions required for the consistency and asymptotic normality of the estimators.

1.2 Some definitions

The sampling model used in this thesis is two-stage sampling, in particular, at the first stage, clusters are selected at random, and then within each cluster, a number of units is selected at random. It is assumed that clusters are distributed independently of each other, but that units within a cluster are correlated with each other, the amount of correlation being measured by a coefficient denoted as the intracluster or intraclass correlation.

The outcome, response or dependent variable will by denoted by Y, and the corresponding explanatory variables or covariates will usually be denoted by X. Although the covariates are sometimes known as independent variables, in this thesis the correlation structure of the distribution of the covariates will be considered, so that in a distributional sense, the X's are not independent, and they will not be referred to as independent variables.

The distribution of the Y's, usually conditional on the X's, will usually be denoted by f(.), and the marginal distribution of the X's will usually be denoted by p(.).

It will be assumed that a generalised linear relationship exists between the Y's and the observed values of the X's, denoted, for example by

$$\Pr(Y=y/x) = g(y/x;\alpha,\beta)$$

where α and β are the parameters of the conditional distribution of Y, given x, in particular, β is the vector of coefficients of the x's, and α is a vector of constants, usually defining the correlation structure of the Y's within a cluster. The intraclass correlation is a residual correlation which exists after all the 'regression' relationships between y and x have been considered.

Throughout the text of this thesis bold notation will usually be used to denote vectors, for example, $\mathbf{x}$, $\boldsymbol{\alpha}$, etc. However, in situations where the use of such notation adds nothing to the discussion, as in the preceding paragraph, or where vectors could not be confused with scalars, the bold notation will not be used and the usual notation will be used, for example, x, α, etc.

1.3 Development of the logistic model

The logistic model was introduced to the statistical literature by Berkson (1944), although he acknowledged its previous use in publications in other fields, in particular, that of physical chemistry. Berkson proposed the following model for bioassay. Let the random variable Y represent death(observed value of 0) or survival(observed value of 1) and let Pr(Y=0) be defined as q where

$$q = 1/[1 + \exp(\alpha - \beta x)] \qquad (1.3.1)$$

where x is the log dosage of a drug or other treatment. Further Berkson assumed that, in an experiment, a small number of different dosages were given and that there were many observations at each dosage level.

In the papers that followed Berkson's, authors debated the relative advantages of different methods of estimation, whether maximum likelihood (Anscombe(1956), Hodges(1958)), minimum chi-square, or Berkson's own minimum logit chi-square (Berkson(1944, 1953, 1955)).

With the maximum likelihood approach, one assumes that, at the i'th dosage level, the number of deaths is the realisation of a binomial random variable with parameters n_i and q_i.

The minimum chi-square estimator, Q_i, minimizes the weighted sum of squares

$$\sum_{i=1}^{k} w_i (q_i - Q_i)^2$$

where k is the number of dosage levels, q_i is the observed failure (mortality) rate at the i'th dosage level, Q_i is the estimated mortality rate (using the model specified in (1.3.1)) at the i'th level, and w_i is the weight for the i'th level, taken to be inversely proportional to the estimated variance of observation q_i, hence

$$w_i = n_i / (Q_i P_i)$$

where

$$P_i = 1 - Q_i$$

and n_i is the number of observations at the i'th level. The minimum chi-square method of estimation is a weighted least squares approach, and, like the maximum likelihood method, is iterative.

The minimum logit chi-square is a non-iterative approach; in particular, it is defined as the solution to the equations

$$\sum_{i=1}^{k} n_i p_i q_i (l_i - m_i) = 0$$

$$\sum_{i=1}^{k} n_i p_i q_i x_i (l_i - m_i) = 0$$

where l_i is the empirical logit transformation

$$l_i = \log(p_i / q_i) \qquad (1.3.2)$$

and m_i is the estimated logit under the model (see 1.3.1), that is,

$$m_i = \alpha - \beta x_i,$$

where x_i is the observed log dosage at the i'th level.

Although the three methods are asymptotically equivalent and efficient, Berkson(1955) showed that, in some cases, the finite sample properties of the minimum logit chi-square estimator are superior to the other two. However, he was forced to conclude(1968, p.85) that 'In some situations, the minimum logit chi-square estimate has the smaller mean square error, in others the maximum likelihood has the smaller... , the difference is minute, even for moderate size samples.'

A problem with the minimum logit chi-square estimator that does not occur with either of the maximum likelihood or minimum chi-square estimators is that it is not defined for cells with all or no failures; hence an estimate cannot be obtained if most cells (dosage levels) have only one observation.

The second step in the development of the logistic model was the suggestion by Dyke and Patterson(1952) of a linear combination of parameters, that is,

$$m_i = \boldsymbol{\beta}'\mathbf{x}_i$$

where the elements of the vector $\boldsymbol{\beta}$ are the main effects and interactions of treatments arranged in a factorial design and $\mathbf{x}_i$ is a vector of variables indicating the presence or absence of each main effect or interaction in cell i of the design. Estimates are obtained by minimizing the distance between these logits and the empirical logits (see 1.3.2). Dyke and Patterson recommended maximum likelihood estimation, but continued to assume that there were many observations at each combination of treatment levels. Grizzle(1961) extended the use of the logistic to more complex factorial designs, and used observational, rather than experimental, studies for his analyses, in particular, those studies in which the data can be described by complex contingency tables. Thus the data that Grizzle analysed usually consisted of non-empty cells with more than one entry per cell. This problem was further examined by Goodman(1963a,b), Birch(1964), and many others.

The third step in the development of the logistic model was taken by Cox(1966) when he wrote the general model

$$\log (p/q) = \boldsymbol{\beta}'\mathbf{x}, \qquad (1.3.3)$$

where

$$\mathbf{x}' = (x_1, \ldots, x_n),$$

and the x_i's can take any value, either the observed value of discrete or continuous random variables, or pre-determined values representing a level of a treatment, or a mixture of both. This model becomes, as a particular case, the complex contingency table model of Grizzle et al. described above. Moreover, Day and Kerridge(1967), using (1.3.3) as a model for discriminant analysis, showed that the distribution of X could be chosen "with complete freedom" (p.314), and Anderson(1972) stated that the distribution of X could be a mixture of continuous and discrete variables.

Mantel(1966) has also discussed a logistic model with 'multiple regressions' (p.92), but failed to write the model explicitly, although he likely meant (1.3.3).

Walker and Duncan(1967) used Cox's model(1.3.3) to analyse data and provided an iterative weighted least squares algorithm to produce estimates.

Truett, Cornfield and Kannel(1967) and Halperin, Blackwelder and Verter(1971) investigated the properties of multivariate normal discriminant functions which can be written as (1.3.3) except that $\boldsymbol{\beta}$, the vector of regression parameters, is constrained to be a predefined function of the means, variances and covariances of the underlying multivariate normal distributions of $\mathbf{X}$ (see Anderson(1958), Cox(1966, p64)). Halperin, Blackwelder and Verter showed that the estimators based on this type of discriminant function behaved poorly if the distribution of X was not a multivariate normal distribution, but rather some discrete distribution, thus showing the need for a more general logistic discriminant function, such as the one proposed by Cox.

The next step in model development had been taken by both Cox(1966) and Mantel(1966) in extending the dichtomous or binary logistic model(1.3.3) to the case of polychotomous or multinomial outcomes. Cox(p.65) suggested the model

$$\begin{aligned} \log \Pr(Y=0) &= -\log d \\ \log \Pr(Y=1) &= \alpha_1 + \boldsymbol{\beta}_1'\mathbf{x} - \log d, \qquad (1.3.4) \\ \log \Pr(Y=2) &= \alpha_2 + \boldsymbol{\beta}_2'\mathbf{x} - \log d, \end{aligned}$$

where

$$d = 1 + \exp(\alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}) + \exp(\alpha_2 + \boldsymbol{\beta}_2'\mathbf{x}).$$

Note that the categories of Y are unordered and are numbered 0,1 and 2 only for convenience. Mantel proposed a similar model but included a third set of parameters, α_0 and β_0. However, in order to obtain estimates, one has to impose identifiability constraints (see Press(1972), p268-272). One such set of constraints

$$\alpha_0 = 0,$$
$$\beta_0 = 0$$

yields Cox's model(1.3.4). Cox's model provides an interesting interpretation of the parameters in the sense that, since

$$\log [\Pr(Y=j)/\Pr(Y=0)] = \alpha_j + \boldsymbol{\beta}_j'\mathbf{x}, \qquad j=1,2,$$

the multinomial model contains two binary logistic models and the coefficient β_{ji} may then be interpreted as the change in the log odds of Y=j in favour of Y=0 for unit change in variable x_i, allowing for all the other covariates.

A further generalisation was also suggested by both Cox(1966) and Mantel(1966). This multivariate logistic model allowed for two correlated binary responses from the same unit on which covariates have also been measured. Cox was again more specific about the form of the model and wrote

$$\begin{aligned}
\log \Pr(Y_1 = 0, Y_2 = 0) &= - \log d, \\
\log \Pr(Y_1 = 0, Y_2 = 1) &= \alpha_{01} + \boldsymbol{\beta}_{01}'\mathbf{x} - \log d, \qquad (1.3.5) \\
\log \Pr(Y_1 = 1, Y_2 = 0) &= \alpha_{10} + \boldsymbol{\beta}_{10}'\mathbf{x} - \log d, \\
\log \Pr(Y_1 = 1, Y_2 = 1) &= \alpha_{11} + \boldsymbol{\beta}_{11}'\mathbf{x} - \log d,
\end{aligned}$$

where d is the appropriate normalising constant.

Mantel went further to say that the α parameters could be expressed as contingency table parameters, thus mixing log-linear and logistic models. Nerlove and Press(1973) studied Mantel's version of the multivariate logistic, writing it such that, for a response cell $(i_1,\ldots,i_q)$ in a q-dimensional contingency table, the probability of an occurrence in this cell may be written

$$\Pr(Y_1 = i_1,\ldots,Y_q = i_q) = \exp(\theta_{i_1,\ldots,i_q}) \Big/ \sum_{I_1} \cdots \sum_{I_q} \exp(\theta_{i_1,\ldots,i_q}) \qquad (1.3.6)$$

where I_j is the index set for i_j, that is, the set of all possible values of i_j. Each parameter θ may be written in terms of anova-type terms, such as

$$\alpha_1,\ldots,\alpha_k$$

$$\beta_1,\ldots,\beta_m$$

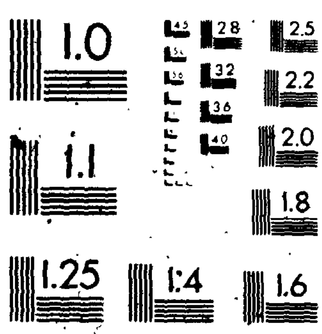
and

$$(\alpha\beta)_{11},\ldots$$

or a regression type of term, such as $\mathbf{c}'\mathbf{x}$, where the x_i's come from continuous distributions, or a mixture of both.

Grizzle(1971) also proposed a multivariate logistic model as a generalisation of the model of Grizzle, Starmer and Koch(1969). However, as a generalisation of that model, it suffers the same drawbacks, that is,

1. it is only defined for discrete values of the covariates, in particular, for each covariate x_i, many observations may occur at


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

(1.3.7) is written in terms of these simple α's. For example, if r of the y's are 1, the rest being 0, the correlated logistic model for that cluster is written

$$\log \Pr(\mathbf{Y}=\mathbf{y}) = \alpha_r + \sum_{i=1}^{n} y_i \boldsymbol{\beta}'\mathbf{x}_i - \log d,$$

where d is the required normalising constant. This model is examined in greater detail in Chapter 2.

A simplified form of the correlated logistic has been studied by Rosner (1983, 1984). In Rosner's model the n α's are constrained so that there are only two α parameters, or alternatively two parameters in an underlying beta or beta-binomial distribution. Rosner suggested that, if c and d are the parameters of a beta distribution, then we may write

$$\alpha_j = \log[\,(c)_j \,/\, (d+n-j)_j\,]$$

where

$$(e)_j = e(e+1)\ldots(e+j-1).$$

For the case with two dependent variables Rosner's model becomes

$$\begin{aligned}
\log \Pr(Y_1 = 0,\ Y_2 = 0) &= -\log d,\\
\log \Pr(Y_1 = 0,\ Y_2 = 1) &= \alpha_1 + \boldsymbol{\beta}'\mathbf{x}_1 - \log d,\\
\log \Pr(Y_1 = 1,\ Y_2 = 0) &= \alpha_1 + \boldsymbol{\beta}'\mathbf{x}_2 - \log d,\\
\log \Pr(Y_1 = 1,\ Y_2 = 1) &= \alpha_2 + \boldsymbol{\beta}'(\mathbf{x}_1 + \mathbf{x}_2) - \log d,
\end{aligned} \qquad (1.3.8)$$

where d is the appropriate normalising constant.

The difference between this simple version of the correlated logistic model and the simple version of Cox's multivariate logisitic model(1.3.5) can be summarised as follows

1. the correlated logistic allows for two values of the covariates, x_1 and x_2, to be associated with each set of values of the dependent variables, (y_1, y_2),

2. the assumption

   $$x_1 = x_2 = x$$

   changes the correlated logistic into a version of the multivariate logistic,

3. the correlated logistic may have some covariates measured at the same level as each y value, that is at the unit level, but may also have some covariates measured at the same level as the pair (y_1, y_2), that is, at the cluster level.

The correlated logistic is proposed for the following types of data

1. data taken on individuals in a sample of families
2. data taken on eyes, ears, hands, etc., in a sample of individuals
3. data arising from repeated measures on the same subject

1.4 Models for clustering

In this thesis it is assumed that sampling is 'by clusters', that is, that there is an infinite, possibly hypothetical, population of clusters (or groups) and that the sample consists of a fixed number (k) of clusters chosen at random; for example, the clusters may be households, families, ears, fingers, etc. An alternative way to explain the sampling method is to refer to it as two-stage sampling, such that first the clusters are chosen at random, and then, within each cluster, the units are chosen at random (from an infinite sub-population of units). An example of this kind of sampling is repeated measures on the same subject. In this thesis, all sampling, whether of families, eyes, or repeated measures will be modelled as two-stage sampling.

The effect of initially sampling clusters is that the values of a variable measured within a cluster tend to be more related to each other than to the values measured on units outside that cluster. In observational data, particularly in family data, the scores within a cluster are usually more positively correlated with each other than they are with scores outside that cluster. However, in designed or experimental studies, the treatment levels may be set such that the scores within a cluster are more negatively correlated with each other than they are with scores outside that cluster.

conditions may be written

1. $\log f(y/r)$ is a differentiable function of β,

2. the function $\log f$ is identifiable with respect to β.

3. the maximum likelihood equations have a unique solution.

A condition peculiar to this problem is that

$$\sum_{i=1}^{\infty} s(\delta,\alpha_i)/i^2 < \infty$$

where

$$s(\delta,\alpha_i) = \mathrm{Var}[\log f(y_i/r_i;\beta+\delta)-\log f(y_i/r_i;\beta)].$$

This condition is required for the strong law of large numbers to be used with the sequence

$$\log f(y_i/r_i;\beta+\delta)-\log f(y_i/r_i;\beta)$$

Andersen's proof of the asymptotic normality of the conditional estimator assumes conditions similiar to those required for the unconditional maximum likelihood estimator with the following additions

1. $h(y)$, the function which bounds the third derivative of $\log f$ has a mean and a variance which are continuous functions of the nuisance parameter

$$\rho = \sigma_a^2/(\sigma_a^2 + \sigma_e^2). \qquad (1.4.1.2)$$

Thus the usual F-statistic tests the hypothesis that

$\rho = 0$.

It may be shown that ρ is indeed a correlation coefficient because, if Y_{ij} indicates the response of unit j within cluster i, then

$$\mathrm{corr}(Y_{ij}, Y_{kl}) = \delta_{ik}\,\rho,$$

where δ_{ik} is the Kronecker delta. Another way to examine the effect of intraclass correlation is to consider the variance of a cluster (or group) mean. If the individual response Y_i has variance σ^2 where

$$\sigma^2 = \sigma_a^2 + \sigma_e^2,$$

and the units are independent of each other, then

$$\mathrm{Var}(\bar{Y}) = \sigma^2/n.$$

However, the presence of intracluster correlation may be seen to increase the variance of $\bar{Y}$, that is, it may be shown that

$$\mathrm{Var}(\bar{Y}) = (\sigma^2/n)[1 + (n-1)\rho]$$

Note that the definition (1.4.1.2) of ρ implies that ρ is always non-negative.

A more general approach to modelling within-cluster correlation with continuous outcomes is to introduce a multivariate model such that the vector Y_i of outcomes in cluster i has distribuiton $(\boldsymbol{\theta}_i, \Sigma)$, where $\boldsymbol{\theta}_i$ is a vector of fixed effects parameters; for example, in the completely randomized design,

$$\boldsymbol{\theta}_i = \mu \mathbf{1},$$

where $\mathbf{1}$ is the vector of 1's. The dispersion matrix of Y_i is written as

$$\Sigma = \sigma^2 [(1 - \rho)I + \rho J],$$

where I is the identity matrix and J is the matrix of 1's. We also assume that the clusters are distributed independently of each other. In this case, this common correlation model produces the same correlation structure as the anova-type model, that is,

$$\text{corr}(Y_{ij}, Y_{kl}) = \delta_{ik}\, \rho.$$

Moreover, the common correlation model allows for negative values of ρ, with the lower limit being $-1/(n-1)$. When one adds the assumption that Y_i has a multivariate normal distribution, the distributional results for the usual F-statistics are the same as those obtained for the anova-type model. The common correlation model for continuous outcomes was first introduced by Kempthorne and Tandom(1953).

The common correlation model has been used in an attempt to model discrete but correlated outcomes. In bioassay, this has given rise to the concept of a normally distributed tolerance distribution for dosage (or log concentration), say

$$X \sim N(\mu, \sigma^2)$$

such that the probability that a subject fails (that is, lacks tolerance) for a dosage x_0 is written as

$$p = \Pr(X < x_0) = F((x_0 - \mu)/\sigma) \quad (1.4.1.3)$$

where F(.) is the cumulative distribution function of the Normal disribution. Hence if n independent subjects are given dosage x_0, the probability that exactly r of them fail is

$${}_nC_r\, p^r\, (1-p)^{n-r}.$$

Similarly if a proportion f of subjects fail at dosage x_0, then the normal equivalent deviate (Gaddum(1933)) or probit (Bliss(1934)) is

$$y = F^{-1}(f).$$

The probit transformation transforms the non-linear relationship of proportion versus dose to a linear relationship of probit versus dose (assuming the underlying normal tolerance distribution is correct). Given data from a bioassay or quantal response, one transforms the proportions $f_1,\ldots,f_k$ to yields probits $y_1,\ldots y_k$ (just as in Section 1.3 one would have transformed the proportions to logits) and fits a straight line of the probits against dosage $x_1,\ldots,x_k$. This methodology is described in Finney(1971).

The probit model was extended to a model for clusters by Ashford and Sowden(1970) who suggested an underlying multivariate normal model for the joint tolerances of the units in a cluster, that is,

$$X \sim MVN(\mathbf{0},\Sigma).$$

As a special case, Σ could be defined as in the common correlation model, and the probability of both unit i and j in a cluster failing at dosage x could be written

$$P_{11}(x) = F(x,x;\rho)$$

where $F(.,.;\rho)$ is the cumulative distribution function of the bivariate normal distribution. Moreover, let

$$P_{10}(x) = P_i(x) - P_{11}(x)$$

$$P_{01}(x) = P_j(x) - P_{11}(x)$$

and

$$P_{00}(x) = 1 - P_{01}(x) - P_{10}(x) - P_{11}(x)$$

where

$P_{kl}(x) = \Pr(Y_i = k,\ Y_j = 1\ /X=x)$

and

$P_k(x) = \Pr(Y_k = 1/X=x)$.

Given this definition of probabilities of the joint occurrence (or failure) distribution of Y, one may estimate the parameters measuring the effect of a single covariate (dosage) on the outcome. The generalisation of this model to more than one covariate will be discussed in Section 1.5.

1.4.2 Beta-binomial models

The presence of intraclass (or intracluster correlation) affects the variance of binary outcomes in a manner analogous to that for continuous outcomes. For example, it may be shown that variance of a sample proportion (for sample size n) increases from pq/n to

$(pq/n)\ [1 + (n-1)\rho]$.

This 'heterogeneity of variance' was recognized in the early 1940's by Cochran(1943). Finney (1971, section 4.6) suggested a 'heterogeneity factor' to be used to adjust the usual variance estimate (in his case, in probit analysis) in order to obtain a better estimate of the variance of an estimator; in particular, he proposed

$h = K/(k - 2)$

where k is the number of clusters, and K is the usual Pearson chi-square, that is,

$$K = \sum_{i=1}^{k} [(r_i - n P_i)^2 / (n P_i Q_i)] .$$

where r_i, P_i and Q_i are the number of occurrences, the estimated probability of occurrence and the estimated probability of non-occurrence, respectively, in cluster i.

Another way in which to consider this extra-binomial variation is to see it as due to some variability in the parameters of the binomial distribution. For example, if p has mean μ and variance σ^2, then, as suggested by Kleinman(1973),

$$\begin{aligned} E(\bar{P}) &= E_P \; E_{\bar{P}/P} \; (\bar{P}) \\ &= E_P \; (P) \\ &= \mu \end{aligned}$$

and

$$\begin{aligned} Var(\bar{P}) &= E_P \; Var_{\bar{P}/P} \; (\bar{R}) + Var_P \; E_{\bar{P}/P} \; (\bar{P}) \\ &= E_P \; [P(1-P)/n] \; + Var_P \; (P) \\ &= [E(P) - E(P^2)]/n \; + \sigma^2 \\ &= \{\mu - [Var(P) + E(P)^2]\}/n + \sigma^2 \\ &= \mu(1-\mu)/n + \sigma^2 \; (n-1)/n. \end{aligned}$$

Both Kleinman and Ishii and Hayakawa(1960) suggested that a good choice for the distribution of P was the beta distribution. When P has a beta distribution, then R, where

$$R = n \; \bar{P},$$

has a beta-binomial distribution. This distribution was investigated by Skellam(1948) and was rediscovered by Ishii and Hayakawa(1960).

The beta-binomial is a compound distribution in which the parameter of the binomial is allowed to take values from a beta distribution with parameters c and d, that is

$$f(p) = p^{c-1} (1-p)^{d-1} / B(c,d), \quad 0 < p < 1,$$

where B(c,d) is a normalising constant known as the beta function. The distribution of R is given by integrating p out of the joint distribution of R and P, yielding

$$f(r) = {}_nC_r \, B(c+r,d+n-r)/B(c,d), \quad r=0,\ldots,n.$$

The probability mass function of the beta-binomial may be written in terms of other functions, for example, the gamma function, so that

$$f(r) = {}_nC_r \, \Gamma(c+r)\Gamma(d+n-r)\Gamma(c+d)/[\Gamma(c)\Gamma(d)\Gamma(c+d+n)]$$

or in terms of

$$(e)_j = e(e+1)\ldots(e+j-1)$$

so that

$$f(r) = {}_nC_r \, (c)_r (d)_{n-r}/(c+d)_n$$

Ishii and Hayakawa(1960) generalised the beta-binomial to the case of polychotomous outcomes, producing the multinomial-dirichlet distribution (under the name of compound-multinomial). Mossiman(1962) is usually given the credit for the latter generalisation. However, Mossiman was the first to realize that there was correlation between the individual Bernoulli-type variates that make up the beta-binomial variable R.

The beta-binomial distribution was used to fit many types of data, for example, by Chatfield and Goodhardt(1970) and Griffiths(1973), the latter proposing a particularly efficient algorithm for maximum likelihood estimation. None of these authors realized that they were making an assumption about the correlations between responses.

Kleinman extended the use of the beta-binomial to the problem of estimation and hypotheses with independent samples, using a weighted least squares algorithm for estimation and partitioning sums of squares to produce an anova table for hypothesis-testing. Williams(1975) used a beta-binomial to analyse a completely randomized design; he used maximum likelihood for estimation and likelihood ratio tests for testing hypotheses about means and variances. Crowder(1978) suggested that similar likelihood methods be used for estimation and testing in general factorial designs, and provided an example for a 2×2 design.

Crowder(1979) was the first to suggest that the correlation between the Bernoulli-type variables be called an intraclass correlation and that it is responsible for the extraneous variance in the cluster means. He wrote the parameters of the beta-binomial as n, μ and ρ, where n is the cluster size,

$$\mu = c/(c+d)$$

and

$$\rho = 1/(c+d+1). \qquad (1.4.2.1)$$

Moreover, he introduced methods for marginal inference about the intraclass correlation.

1.4.3 Other simple models

Cohen(1976) and Altham(1976) proposed models for binary outcomes which allowed for dependence between the individual units yet which required only the first two moments of the joint multivariate distribution. Cohen's model handles only two units per cluster and two possible responses. It is described in detail in Appendix A. Altham's model is a generalisation of Cohen's and handles n units per cluster and possibly more than two response categories; for example, with n units per cluster and 2 outcome categories, and the same marginal probability distribution for each unit, that is

$p = Pr(Y_i = 1),\ i=1,\ldots,n,$

Altham's model yields the following expression for the probability of r 1's and (n-r) 0's

$$
\begin{aligned}
Pr(R=r) &= \rho p + (1-\rho)\, p^n, && \text{if } r=n, \qquad (1.4.3.1)\\
&= \rho(1-p) + (1-\rho)(1-p)^n, && \text{if } r=0,\\
&= (1-\rho)\; {}_nC_r\; p^r\, (1-p)^{n-r}, && \text{otherwise},
\end{aligned}
$$

where ρ is the intraclass correlation $(0 \leq \rho < 1)$. However, Donald(1984) has shown that the probability distributon defined by Cohen-Altham has certain difficulties, in particular, the effect of the clustering or correlation parameter ρ is too weak to overcome the effect of the other parameters. One intuitively expects that for a large enough value of ρ the probability of n-1 1's will exceed the probability of n-2 1's, but, from (1.4.3.1) it follows that

$$\Pr(R=n-2) = (n-1)\,[\Pr(R=n-1)]/2.$$

In fact ρ does not appear in the relationship at all.

Another alternative model to the beta-binomial is the correlated binomial developed by Kupper and Haseman(1978). Bahadur(1961) proposed a general model for correlated binary outcomes such that

$$\Pr(R=r) = P_1(r) \times f(y_1, \ldots y_n),$$

where $P_1(r)$ is the probability under the usual(uncorrelated) binomial model and $f(.)$ is a correction for the lack of independence between the Y's. Bahadur has shown that, if Y_i is standardised so that

$$Z_i = (Y_i - p)/\surd[p(1-p)],$$

then $f(.)$ may be written

$$1 + \sum_{i<j} E(Z_i Z_j)\, z_i z_j + \sum_{i<j<k} E(Z_i Z_j Z_k)\, z_i z_j z_k + \ldots$$

$$+ E(Z_1 \ldots Z_n)\, z_1 \ldots z_n$$

which is a function of the second to n'th order correlations among the Z's and hence among the Y's. In the correlated binomial, only the first order correlation is kept, that is, the probability mass function is written

$$\Pr(R=r) = P_1(r)[1 + \sum_{i<j} E(Z_i Z_j)\, z_i z_j]. \qquad (1.4.3\ 2)$$

Further, since the correlation structure of the Y's is symmetrical, the correlation between any pair of Y's is the same and equal to ρ. Moreover, since all y's are 0 or 1, all z's are equal to $-p/s$ or to $(1-p)/s$, where

$$s = \surd[p(1-p)],$$

the probability in (1.4.3.2) may be written as

$$P_1(r)\ \{1 + \rho[p^2(n-r-1)(n-r-2) - p(1-p)r(n-r) + (1-p)^2 r(n-r)]/s^2\}, \text{ for } r > 1,\ n-r > 1. \qquad (1.4.3.3)$$

A limitation of the correlated binomial is that the mass function is only non-negative for a limited range of values of r. This range of values decreases with increasing cluster size and as p moves away from 0.5; for example, for cluster size of 5 and p value of 0.1, the range of possible values is only (-0.01,0.30). For further ranges, see Kupper and Haseman(1978, p.72).

Beacuse of the problems described above, the Cohen-Altham and correlated binomial models are not considered appropriate for further study in this thesis.

1.5 Clustering models with covariates

1.5.1 Models based on normally distributed error

Pierce and Sands(1975) proposed a model to handle extra-binomial variation in the presence of covariates. The relationship between the binary outcome and the sum of the explanatory (or systematic) component and the error (or random) component is modelled by a logistic transformation; that is, if, at the value of the covariate for a cluster there are the n responses of the units within that cluster, and, if r of the responses are positive then we have a binomial model, namely,

$$Pr(R=r) = {}_nC_r \; p^r \; (1-p)^{n-r}$$

where,

$$\text{logit } p = \boldsymbol{\beta}'\mathbf{x} + z,$$

where the random (or error) component, z, has been added to the systematic component (or covariate effect). Pierce and Sands next assumed that

$$Z \sim N(0,\sigma^2),$$

for some unknown σ^2. Further Z is assumed to be independently distributed of X and of all other Z's. This logistic-normal model may be used only for cluster-specfic covariates. Moreover, estimation is tedious because of the numerical integration required for maximum likelihood estimation. Pierce and Sands proposed instead a two-stage method, first estimating σ^2, then using weighted least squares for estimating $\boldsymbol{\beta}$.

Another model in this class of linear-normal models is the correlated probit regression model of Ochi and Prentice(1984), in which one assumes, at each value of x, the equicorrelated multivariate probit model of Ashford and Sowden, where 'equicorrelated' implies that the covariance matrix is that of the common correlation model of section 1.4.1. Ochi and Prentice proposed the use of maximum likelihood methods for estimation where numerical integration is accomplished by the numerical approximation of the equicorrelated multivariate cumulative distribution function in terms of the univariate normal cumulative distribution function. This model also was proposed only for cluster-specific covariates.

Korn and Whittemore(1979) also produced a logistic-normal model, but appropriate only for unit-specific variables. Let

$$\mathrm{logit}(p_{ij}) = \beta_i' x_{ij},$$

where p_{ij} is the theoretical probability for the j'th unit in cluster i, x_{ij} is the value of its covariate, and β_i is the parameter for that cluster. A random component was introduced (equivalent to introducing intraclass correlation) by assuming that

$$\beta_i \sim MVN(\beta, D) \qquad (4.5.1.1)$$

where now the vector β and the matrix D are the unknown parameters. This model has a two-stage solution,

1. estimate the β_i for each cluster using the usual maximum likelihood estimator for logistic models.

2. determine the posterior distribution of β and D given the β_i's and their variances and obtain estimators which maximize these posterior distributions. In particular, the estimate of the common β is merely a weighted average of the β_i's (see Korn and Whittemore(1979), p.798).

Laird and Ware(1981) and Stiratelli, Laird and Ware(1984) introduced a linear-normal model that can handle both cluster- and unit-specific covariates and allow for correlation between binary outcome variables. In this model we have

$$\text{logit}(p_{ij}) = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}_i'\mathbf{z}_{ij}$$

where p_{ij} and $\boldsymbol{\beta}$ are the same as for the Korn-Whittemore model, $\mathbf{x}_{ij}$ is the vector of covariates for the j'th unit of the i'th cluster $\boldsymbol{\gamma}_i$ is the parameter related to the random component, and $\mathbf{z}_{ij}$ is a vector of indicator variables. This is also a two-stage model, and at the second stage, the authors assume that

$$\tilde{\boldsymbol{\gamma}}_i \sim MVN(\mathbf{0}, D)$$

and, in order to use a Empirical Bayes approach for estimation, that $\boldsymbol{\beta}$ has a diffuse prior distribution. It is not clear whether the cluster-specific covariates are part of the x vector or are only related to it by some linear transformation. Neither the text of the paper nor the example indicate how the cluster-specific covariates are included.

The authors proposed a restricted maximum likelihood estimation (REML) of the variance components of D which is equivalent to integrating out the fixed effects over a diffuse prior in an empirical Bayes approach. Integration over many dimensions is avoided by using posterior modes instead of means. Finally the EM algorithm (Dempster, Laird and Rubin(1977)) is used to obtain the estimates.

1.5.2 Other models with covariates

Manly(1978) and Williams(1982) proposed that the extra-binomial variation in a linear logistic model be handled in a method similar to that proposed by Kleinman(1973) (see Section 1.4.3); that is, let the parameter μ be related to the covariates by the logistic transform

$$\text{logit}(\mu) = \boldsymbol{\beta}'\mathbf{x},$$

and the extra variation be induced by the correlation between the components $Y_1,\ldots,Y_n$, where

$$E(Y_i) = \mu,$$

$$Var(Y_i) = \mu(1-\mu),\ i=1,\ldots,n.$$

If we write

$$Corr(Y_i,Y_j) = \rho$$

and, if,

$$R = \sum_{i=1}^{n} Y_i$$

then R has a distribution such that

$$E(R) = n\mu$$

$$Var(R) = n\mu(1-\mu)[1+\rho(n-1)]. \qquad (1.5.2.1)$$

Manly focussed on data fitted by a special exponential distribution while Williams concentrated on the beta-binomial (see section 1.4.2). However, both were also interested in quasi-likelihood methods which can be used for estimation when only the first two moments of R are known. Williams produced estimates according to the following algorithm:

1. for known ρ (initially 0) the partially specified model can be fitted as a generalised linear model using the program GLIM(Nelder and Wedderburn(1972), Baker and Nelder(1978), McCullagh and Nelder(1984)).

2. successive values of ρ are obtained by setting the goodness-of-fit statistic equal to its expected value.

This two-step procedure is repeated until convergence is obtained for the estimate of ρ. The maximum likelihood estimates of β are given by the appropriate values from step 1. Note that the Manly-Williams model includes only cluster-specific covariates.

Rosner(1983, 1984) proposed a model which is a generalisation of the Williams model in that it has a beta-binomial error structure, but can handle both cluster- and unit-specific covariates. This model was introduced in Section 1.3 and is covered in full detail in Chapter 2.

1.6 Related methods

1.6.1 Dummy variables (stratified analyses)

The use of dummy variables to represent treatments (that is, levels of a factor) allows the analysis of variance or the analysis of covariance to be performed by means of a regression model. A particular dummy or indicator variable is assigned a value of 1 for those experimental units which receive the level of the factor represented by the dummy variable and 0 for all other units (see Draper and Smith(1981, p.241-50), Kleinbaum and Kupper(1978, p.256-9)). This methodology is also used in observational studies to handle the modelling of covariates, such as sex, socio-economic status, etc., that take discrete values.

It must be noted that this method provides optimal estimators and tests when the factor or variable being modelled represents a fixed (as opposed to a random) effect. However, in a completely randomized design, the dummy variable regression approach does provide a test statistic which is correct for both the fixed and the random effects models. However the test statistic does have a different distribution under the alternative hypothesis specified by the two different models. Moreover, for even slightly more complicated designs, such as a two-way factorial, or a one-way anaylsis with a single covariate, the dummy variables approach does not provide the correct statistic for the random effects model. Hence the dummy variables approach only provides an optimal test statistic for a random effects model in a very special case (see also Winer(1971), p.220-8, p.244-51).

One case in which the dummy variables aproach (usually called stratified analysis) would seem to be appropriate occurs in the analysis of a case-control study in which the population is considered to be divided into k strata on the basis of the matching variables (Breslow, Day, Halvorsen, Prentice and Sabai(1978)). The outcome variable measures occurrence(case) or non-occurrence(control) of a disease, and a dummy variable is used to indicate membership of each stratum. The model may be written

$$\text{logit}(p_{ij}) = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{ij}, \quad (1.6.1.1)$$

where p_{ij} is the theoretical probability that unit j in cluster(stratum) i has value 1 (is a case), $\mathbf{x}_{ij}$ is the value of the covariate for that unit, and α_i is the coefficient of the dummy variable (say v_{ij}).

Since the stratum membership is a fixed effect, the dummy variables method would seem to be appropriate. However, the maximum likelihood estimator of the parameters is only unbiased asymptotically (in n, the stratum size). The bias for small strata is large (Andersen(1973), Pike, Hill and Smith(1980), Chamberlain(1980), Lubin(1981), see also Section 1.7.3). A cohort study may also be stratified on the basis of values of covariates and analysed using the stratified (dummy variables) model(1.6.1.1) (see Pike, Hill and Smith(1980).

1.6.2 Conditional models

An alternative procedure to stratified (dummy variables) analysis in case-control studies or, equivalently, in certain cohort studies, is, in each stratum, to condition on the number of cases and controls, or, in the cohort study, on the number of occurrences and non-occurrences. The conditional model may be written

$$\frac{\prod_{i=1}^{r} \exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{j=1}^{{}_nC_r} \prod_{i\in I_j} \exp(\boldsymbol{\beta}'\mathbf{x}_i)} \qquad (1.6.2.1)$$

where r is the number of cases(occurrences), and I_j is the index set of the cases. This conditional method was used by Liddell, McDonald and Thomas(1977), Breslow, Day et al.(1978), and Prentice and Breslow(1978), but is usually referred to as the Breslow-Day model.

It will be shown in Chapter 2 that the correlated logistic model (see Section 1.3), under the same conditioning described above, becomes a conditional model equivalent to the Breslow-Day model described in (1.6.2.1).

1.7 Inference under incorrect models

The purpose of this thesis is to introduce a model for correlated binary outcomes in the presence of covariates, namely, the correlated logistic regression model, and to examine the behaviour of certain estimators under this model, in particular, to compare the asymptotic means and variances of estimators which are calculated without using the probability structure of the correlated logistic, and to investigate how tests based on non-optimal estimators might be adjusted to improve their large-sample behaviour.

Similar comparisons of estimators have been made in three areas

1. the behaviour of Pearson chi-square tests for contingency table analysis when the underlying probability distributions display clustering.

2. the behaviour of ordinary least squares estimators when the common correlation model holds.

3. the behaviour of unconditional and conditional estimators in matched case-control studies and in stratified cohort studies.

This literature will be reviewed briefly below.

1.7.1 <u>Contingency table analysis</u>

Both Cohen(1976) and Altham(1976), in discussing their models for clustered binary responses measured the effect of the intraclass correlation, by demonstrating how it changed the distribution of the usual Pearson chi-square statistic for testing independence in an RxC contingency table, X^2. The limiting distribution of X^2 under independence of units is a chi-square on (R-1)(C-1) degrees of freedom, but under the Cohen model with 2 (correlated) units per cluster,

$$X^2 \sim (1+\rho)\, \chi^2_{(R-1)(C-1)},$$

where ρ is the intracluster correlation. Similarly Altham showed that, for clusters of size n,

$$X^2 \sim [1 + (n-1)\rho]\, \chi^2_{(R-1)(C-1)}.$$

Brier (1980) introduced the dirichlet-multinomial as a model for more than two (say r) outcome categories and for n units per cluster. He extended Altham's results to show that, for the hypothesis test,

$$H_o: p_i = f(\theta)$$

then both the Pearson chi-square statistic and the likelihood ratio chi-square statistic G^2 have the asymptotic distribution

$$C \chi^2_{r-s-1},$$

where s is the dimension of the reduced parameter space under H_o, and

$$C = [1 + (n-1)\rho]$$

Moreover, for a set of k clusters of varying size $(n_1, \ldots n_k)$, such that

$$\sum_{i=1}^{k} n_i = N$$

both X^2 and G^2 have the asymptotic distribution

$$B \chi^2_{r-s-1},$$

where

$$B = \sum_{i=1}^{k} n_i C_i / N$$

and

$$C_i = 1 + (n_i - 1)\rho.$$

Rao and Scott(1981) examined the distribution of the Pearson chi-square statistics under complex survey designs and showed that both the chi-square for goodness of fit, X^2, and the chi-square for independence, X^2_I, are asymptotically distributed as weighted sums of χ^2_1 variables, where the weights are related to deffs, that is, design effects (the variance under a specific sampling design as compared to simple random sampling). Rao and Scott state that the

Cohen, Altham and Brier models all correspond to the assumption of a constant design effect, thus yielding exact correction factors for the chi-square statistic. Under the assumption of multiple design effects, Rao and Scott obtained conservative corrections to X^2 and X_I^2. Rao and Scott(1983) extended these results to multiway tables.

Thomas and Rao(1984) extended Brier's model to generate non-constant design effects by allowing a mixture of dirichlet-multinomial probability vectors for each cluster, but retaining the same intracluster correlation, and investigated the behaviour of various adjustments to the Pearson and log-likelihood ratio chi-squares statistics.

1.7.2 Least squares estimators

Some researchers have investigated the properties of least squares estimators when certain assumptions about the model are incorrect; in particular, when

$$\mathbf{Y} \sim MVN(\boldsymbol{\beta}'X, \sigma^2 \Sigma)$$

but estimation assumes that

$$Var(\mathbf{Y}) = \sigma^2 G$$

where

$$G \neq \Sigma .$$

Watson(1955) examined the problem for

$$G = I$$

and obtained bounds for the bias of the least squares estimator, a lower bound for the efficiency of the estimators, and bounds on the significance points of t- and F-tests. He based his definition of relative efficiency on the ratio of generalised variances, that is,

$$|var(\hat{\beta}_1)| / |var(\hat{\beta}_2)|, \qquad (1.7.2.1)$$

where $\hat{\beta}_1$ is the generalised least squares estimator and $\hat{\beta}_2$ is the ordinary least squares estimator. Watson(1967) dropped the assumption of underlying multivariate normal distributon and proposed a lower bound on (1.7.2.1), but was unable to prove it. Knott(1975) and Bloomfield and Watson(1975) both obtained the result that the (1.7.2.1) is bounded below by

$$\prod_{i=1}^{\min(p,n-p)} [4 \gamma_i \gamma_{n-i+1} / (\gamma_i + \gamma_{n-i+1})^2]$$

where p is the number of covariates and n is the number of units in the sample and

$$\gamma_1 > \dots > \gamma_n > 0,$$

where $(\gamma_1, \dots, \gamma_n)$ are the eigenvalues of Σ.

This result only yields a lower bound. Sharper bounds have been obtained, in particular, McElroy(1967) and Kruskal(1968) have shown that, when ρ is known, the relative efficiency of the ordinary least squares estimator is exactly 1 if and only if

$$\Sigma = \sigma^2 [(1-\rho)I + \rho J],$$

where I is the identity matrix and J is th matrix of 1's, that is, when Σ is the common correlation matrix of Section 1.4.1. In this case the ordinary least squares estimator is equivalent to the generalised least squares estimator which is the best linear unbiased estimator.

McElroy further showed that, when the common correlation model holds, most elements of the covariance matrix of the ordinary least squares estimator are estimated unbiasedly by $s^2(X'X)^{-1}$, where s^2 is the usual estimator of σ^2. Only the elements of the first row and column are not estimated correctly; for example, the estimated variance of β_0 is wrong.

Halperin(1951) showed that certain estimators and tests are equally valid in the case of the common correlation model; in particular, the estimators of β, except for β_0, from the ordinary least squares method, have the required multivariate normal distribution, and the usual sums of squares have independent chi-square distributions and hence their ratios have the required F distributions.

Scott and Holt(1982) investigated how clustering as well as samples of unequal size affect the properties of the least squares estimator. Again, this estimator remains unbiased, but is no longer equivalent to the generalised least squares estimator and suffers some loss of efficiency. An upper bound to the loss of efficiency is given by

$$m^2\rho^2 \;/\{m^2\rho^2 + 4(1-\rho)[1+(m-1)\rho\} \qquad (1.7.2.2)$$

where ρ is the intraclass correlation and m is the maximum cluster size. Since ρ measures the residual correlation after the regression on X, the loss of efficiency is usually small. Moreover Scott and

Holt calculated the misspecification factor, namely the amount which must be used to multiply the usual estimate of the variance-covariance matrix of the least squares estimator in order to get the true variance-covariance matrix of the estimator. In this case the factor is

$$(X'\Sigma X)(X'X)^{-1}$$

$$= I + (M-I)\rho$$

where

$$M = \sum_{i=1}^{k} (n_i X_{B_i}{}'X_{B_i}) (X'X)^{-1},$$

$$X_{B_i} = (\mathbf{x}_i{}'\mathbf{1})\ \mathbf{1},$$

and $\mathbf{x}_i$ is the i'th row of X and $\mathbf{1}$ is the vector of 1's.

Christensen(1984) has extended the work of McElroy to show that, in a model with dummy variables indicating membership of the clusters, the ordinary least squares estimator is the same as the generalised least squares estimator (assuming that ρ is known), both of which are best linear unbiased estimators, and that the covariance of the regression estimates can be calculated in the usual way. However, this optimality exists only when the cluster effects are considered to be fixed effects, and this is usually <u>not</u> the case.

In addition, the effect of sampling designs on the estimation of regression parameters has been studied in some detail by Konijn(1962), Fuller(1975), Nathan and Holt(1980) and Holt, Smith and Winter(1980).

1.7.3 Case-control studies

In case-control studies, under the assumption of equal relative risk in each cluster(stratum) the prospective model(1.6.1.1) is appropriate. It might seem that, since in case-control studies the sampling is retrospective, the proper model would have exposure to a risk factor as the dependent(outcome) variable, as proposed by Prentice(1976). This is known as the retrospective model. Breslow and Powers(1978) have shown that the estimates of the regression coefficients under the two models agree quite closely when the models contain many covariates and the estimates are equivalent when the models are fully saturated. Moreover, Prentice and Pyke(1979) extending some work of Anderson(1972), showed that in cas -control studies (with dummy variables for strata) the estimators under both prospective and retrospective models are equivalent. The prospective model(1.6.1.1) can be written in the same format as the usual logistic model (including, of course the dummy variables) and hence is preferred for numerical calculations. Cohort studies in which strata exist may also be analysed by use of the prospective model.

However, an alternative analysis to the unconditional approach of the prospective model is the use of a conditional model in which the stratum parameters are considered as nuisance parameters and are removed by conditioning (see model (1.6.2.1)).

Pike, Hill and Smith(1980) have compared these two methods of estimation and have shown

1. in small strata, the unconditional analysis is badly biased, whereas the conditional estimators are unbiased.

2. in large strata, the bias of the unconditional estimator becomes small.

Lubin(1981) repeated this investigation but concentrated on small to moderate-sized samples, showing that

1. the unconditional method tends to overestimate the true risk unless the stratum size is large.

2. for matched pairs data the conditional method underestimates the true risk, but for larger strata (say, 3 cases, 3 controls) this underestimation disappears.

Farewell(1979) has shown that, for a single stratum, the conditional estimator is fully efficient when the relative risk is 1, and has high efficiency for other values of the relative risk. He further argued that

1. these results hold exactly for many strata of the same size.

2. for varying strata size, the results should hold (and he gives one example where this is true).

1.8 Asymptotics

This section deals with the assumptions made in this thesis about the probability distribution of the dependent and explanatory variables in order that the estimators produced have the required asymptotic properties.

1.8.1 Consistency and asymptotic unbiasedness

Kendall and Stuart(1973, p.5) state that
'a consistent estimator whose asymptotic distribution has finite mean value must be asymptotically unbiassed'.

In this thesis we shall be examining the behaviour of estimators of the regression coefficient, β, for finite values of β, in particular, for

$$\beta = 0.$$

Hence, by Kendall and Stuart's definition we may use the terms consistency and asymptotic unbiasedness, and lack of consistency and asymptotic bias, interchangeably.

1.3.2 Maximum likelihood estimation

Maximum likelihood estimation produces estimators of a parameter θ from observations, $\{z_i; i=1,\ldots,k\}$, with probability density function

$$f(z_i;\theta)$$

With a sample of size k, it is usually assumed that the observations are independently and identically distributed.

However, for logistic models, the form of the density function varies with the observed value of the covariate X, so that if we write

$$z_i = (y_i, x_i)$$

then, for the i'th observation

$$f_i(y_i, x_i;\theta) = f_i(y_i/x_i;\theta)g(x_i;\theta)$$

where the form of the numerator of $f(y_i/x_i;\theta)$ depends on the value of x_i.

This occurrence of non-identically distributed random variables $\{Y_i; i=1,\ldots\}$ does not allow the use of the usual regularity conditions for the asymptotic properties of maximum likelihood estimators as stated in Wald(1949) and Cramer(1946) and summarised, for example, in Cox and Hinkley(1974, Chapter 9).

Bradley and Gart(1962) suggested an approach to problems like this one in which the observations do not come from a single population but from associated populations which have parameters in common. In the case of correlated logistic models, each value of

$$Z = (Y,X)$$

where Y is the vector of dependent observations in a cluster and X is the matrix of corresponding values of the covariates, defines a population; each populaton shares the common logistic regression parameters, β, and some or all of the constant, or α, parameters.

Bradley and Gart established the sufficient conditions for the distribution of Z such that the usual asymptotic results will hold for the maximum likelihood estimators of the parameter θ, where

$$\theta = (\alpha,\beta)$$

that is, that

$$\theta \sim N(\theta_0, I^{-1})$$

where I is the expected information matrix whose (i,j)'th element is

$$\lim_{k \to \infty} \frac{1}{k} \sum_{h=1}^{k} E\left(\frac{\partial^2 \log f_h}{\partial\theta_i \partial\theta_j} \right)_{\theta=\theta_0} \qquad (1.8.2.1)$$

The conditions are the following:

1. the density functions within each population follow the usual requirements for consistency and asymptotic normality of maximum likelihood estimators (see, for example, Cox and Hinkley(1974, p.281), that is,

1.1 the first three derivatives of the log likelihood with respect to any subset of the parameters exist in a neighbourhood containing the true parameter value almost surely.

1.2 the first and second derivatives of the densities are bounded above by functions of Z which are integrable over the sample space.

1.3 the absolute value of the third derivative of the log likelihood is bounded above by a function whose expectation exists.

1.4 the matrix defined in (1.8.2.1) is positive definite with a finite determinant

2. conditions which constrain the distribution of any of the Z_i's so that they are not too extreme; for example, the probability of obtaining a value of the first derivative of any log likelihood greater than k must be o(1); similarly for the second derivative. For further discussion see Bradley and Gart, p.210-11.

The conditions in 2. allow Bradley and Gart to prove consistency and asymptotic normality using a theorem of Cramer(1947) and Ghanda(1954).

Gourieroux and Montfort(1981) established more restrictive conditions on the distribution of X in the case of the binary (or dichotomous) logistic model, namely,

1. the X's are uniformly bounded.

2. for sample size n, if λ_{1n} and λ_{kn} are smallest and largest eigenvalues, respectively, of

$$\left(\frac{\partial^2 \log f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)_{\theta=\theta_0}$$

the ratio, $\lambda_{1n}/\lambda_{kn}$, is bounded for every n.

The latter condition is implied by the usual condition on the information matrix given in 1.4 above.

Hauck(1982), in dealing with the multinomial (or polychotomous) logistic regression model, suggested an alternative set of conditions, namely,

1. the parameter, θ, can be partitioned into subparameters, θ_1 and θ_2, where θ_1 is the parameter of the conditional distribution of the dependent variables and θ_2 is the parameter of the marginal distribution of the covariates. In this case we can write the joint distribution function as

$$f_i(y_i, x_i; \theta) = f_i(y_i / x_i; \theta_1)\ g(x_i; \theta_2) \qquad (1.8.2.2)$$

2. the joint distribution of the Y's and X's satisfies the usual conditons for consistency of maximum likelihood estimators.

When these conditions hold, the first derivatives of the log likelihood with respect to θ_1 is not a function of θ_2, and visa versa, and, similarly the matrix of second derivatives is block diagonal, so that asymptotically, the estimates of θ_1 are independent of the estimates of θ_2. Thus we may ignore the parameters of the distribution of the covariates except where they appear explicitly in expectations, for example, in the calculation of

$$E\left(-\frac{\partial^2 \log f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right).$$

In this thesis we use Hauck's assumptions with respect to the correlated logistic model. In particular, we assume that the distribution of X is not a function of θ_1 and that the conditional distributon of Y is not a function of θ_2. Otherwise the results described above hold. The only difference is that now each of the associated populations is a cluster and the outcome variable is a vector **Y** and the covariate is described by a matrix X consisting of the n covariate vectors in a cluster.

Next we must consider whether the usual conditions are satisfied by the joint distribution of **Y** and X as specified by the vector version of (1.8.2.2), that is,

$$f(\mathbf{y}_i, X; \theta) = f(\mathbf{y}_i/X; \theta_1)\, g(X; \theta_2) \qquad (1.8.2.3)$$

where $f(\mathbf{y}_i/X;\theta_1)$ is the corralated logistic distribution and $g(X;\theta_2)$ is an unspecified distribution. The first derivative of f with respect to θ_1 where

$$\theta_1 = (\alpha, \beta)$$

is merely f times X_i, an element of the vector random varible **X**. Similarly, as shown in section 3.4.1, the second derivative of f is a product of terms like f times $X_i X_j$ where X_i and X_j are elements of **X**, and the third derivative may be shown to be a product of terms like f times $X_i X_j X_k$. Condition 1.5 requires that we take expectation with respect to the distribution of **X**. Since all terms like f are bounded above by **1**, the condition is satisfied by a distribution of **X** with finite third moment. This condition is not sufficient because we also require that g, the distribution of **X**, satisfy the regularity conditions for the derivatives of g with respect to θ_2, the parameter of the distribution of **X**.

However, we shall assume that the distribution of **X** is such that the regularity conditions for the distribution of **X** are satisfied, as are those for the joint distribution of Y and **X**.

1.8.3 The conditional maximum likelihood estimator

The correlated logistic model may be considered to contain two parameters, the regression coefficient parameters, β, which is of interest, and the constant parameter, α, which may be thought of as a nuisance parameter. It is well-known that, in the presence of nuisance parameters, inferences may be made about a parameter of interest by using a likelihood obtained by conditioning on a statistic, say C. If the statistic C has a marginal distribution which is not a function of β then C is an ancillary statistic for β

(Fisher(1956)), and the conditional likelihood is conditionally sufficient for β (see also Cox and Hinkley(1974, p.32,35)).

Andersen(1970) produced conditional maximum likelihood estimators by introducing a different type of ancillarity. He required that C be a minimal sufficient statistic for α for every value of β, but the distribution of C may be a function of both α and β. However, the distribution of the data, conditional on C, is not a function of α, and hence may be used to produce conditional maximum likelihood estimators.

As will be shown in section 2.4, with the correlated logistic model in a single cluster, the statistic R, which measures the number of occurrences (that is, the number Y's which have value 1), has a distribution which is a function of both α and β. However, the conditional distribution, given R, does not contain α. This may be explained intuitively in the following way: for a given value of β, the value of R indicates the relative propensity of an occurrence, but this propensity is indicated in the model by the parameter α. Hence, given β, R contains all the information about α, so that conditioning on R removes α from the inference problem.

Andersen(1970) showed that, under conditions similar to those required for the unconditional maximum likelihood estimator, the conditional estimator converges almost surely to the true value. In terms of the parameters and variables used in the correlated logistic model, these conditions may be written

1. $\log f(y/r)$ is a differentiable function of β,

2. the function $\log f$ is identifiable with respect to β.

3. the maximum likelihood equations have a unique solution.

A condition peculiar to this problem is that

$$\sum_{i=1}^{\infty} s(\delta,\alpha_i)/i^2 < \infty$$

where

$$s(\delta,\alpha_i) = \mathrm{Var}[\log f(y_i/r_i;\beta+\delta)-\log f(y_i/r_i;\beta)].$$

This condition is required for the strong law of large numbers to be used with the sequence

$$\log f(y_i/r_i;\beta+\delta)-\log f(y_i/r_i;\beta)$$

Andersen's proof of the asymptotic normality of the conditional estimator assumes conditions similiar to those required for the unconditional maximum likelihood estimator with the following additions

1. $h(y)$, the function which bounds the third derivative of $\log f$ has a mean and a variance which are continuous functions of the nuisance parameter

2. f, the mean and the variance of the second derivative are continuous functions of the nuisance parameter.

It can be shown that, for the correlated logistic model, conditional on the observed values of the covariates, Andersen's conditions hold. If we now introduce the distribution of the covariates, as with the unconditional estimator, we require that the first three moments of the distribution of **X** exist and are finite. Under these conditions the consistency and asymptotic normality of the conditional estimator are assured.

1.8.4 Estimators under incorrect models

As will be discussed in several sections of Chapter 3 we are interested in the behaviour of estimators calculated under the wrong models. However, as is shown in section 3.3.3 and in Appendix B.4, these estimators are obtained by using the wrong estimating equations. In determining regularity conditions for the consistency and asymptotic variance of these estimators, we can replace log f in the conditions already discussed above with the particular function used in the wrong estimating equation, say log h. However, as will be shown in this thesis, the derivatives of log h yield expressions similar to those of log f; for example, the second derivative consists of terms like h [= Pr(Y=1)] times $X_i X_j$, so that the requirements are the same as those established for the maximum likelihood estimators in section 1.8.2, that is,

52

1. the distribution of **X** must have finite third moments.

2. the distribution of **X** must satisfy the regularity conditions for estimating the parameters of that distribution.

CHAPTER 2

Characteristics of the Correlated Logistic Model

2.1 Introduction

This chapter describes certain properties of the correlated logistic model which was defined in Section 1.3. The properties discussed are limited to those required to determine the correlation structure of the model and to establish the relationship with other models used in this area, and are not a complete description of this family of models. Section 2.2 covers the general correlated logistic model, including two ways of representing the model and the relationship between these two representations. In addition we examine the first and second moments of the marginal and joint (for 2 variables at a time) distributions of the outcome variables. Section 2.3 discusses the properties of the reduced correlated logistic proposed by Rosner, showing in particular the relationship with a generalised hypergeometric distribution and with the beta-binomial distribution. Section 2.4 covers the conditional correlated logistic model, where conditioning is imposed as suggested by Breslow, Day et al.(1978) for case-control studies. It is shown that the resulting conditional model is the same as that obtained by Breslow and Day(1980).

Note: In this chapter the notation $_nC_j$ is generally used for the number of combinations of n items taken j at a time, that is,

$$_nC_j = \binom{n}{j} = \frac{n!}{j!(n-j)!}$$

2.2 The correlated logistic model

The correlated logistic model for an outcome variable Y may be written

$$\log \Pr(\mathbf{Y}=\mathbf{y}/X) = \alpha_r + \sum_{i=1}^{n} y_i \boldsymbol{\beta}'\mathbf{x}_i - \log d, \quad (2.2.1)$$

where n is the cluster (or group) size, r is the number of 1's in the cluster, that is, the number of elements in Y that are equal to 1, the other n-r elements being equal to 0, $\mathbf{x}_i$ is the i'th column of X, and d is the normalising constant, in particular,

$$d = \sum_{j=0}^{n} \{ \sum_{h=1}^{_nC_j} [\exp(\alpha_j + \sum_{i \in I_h} \beta'\mathbf{x}_i)]\},$$

where I_h is the index set, that is, the set of indices of the y's that have value 1. For each value of j, there are $_nC_j$ index sets.

This model may also be written in terms of the generalised exponential family defined by Dempster(1971, p327-36), namely,

$$\log \Pr(\mathbf{Y}=\mathbf{y}/X) = \boldsymbol{\beta}' \sum_{i=1}^{n} y_i \mathbf{x}_i + \gamma_1 \sum_{i=1} y_i + \gamma_2 \sum_{i<j} y_i y_j$$

$$+ \gamma_3 \sum_{i<j<k} y_i\, y_j\, y_k + \ldots + \gamma_n \prod_{i=1}^{n} y_i - \log d, \quad (2.2.2)$$

where γ_j is a function of α_1 to α_j, for $j=1,\ldots n$.

It is possible to establish the relationship between these two expressions for the probability distribution function, that is, between the α's and the γ's. Let r be the number of y's that are equal to 1. By expression (2.2.1), the correct constant, that is, the term that does not include any value of x, in $Pr(\mathbf{Y}=\mathbf{y}/X)$ is α_r. However, in expression (2.2.2) the constant is

$$\gamma_1 \sum_{I_r} y_i + \gamma_2 \sum_{I_r} y_i\, y_j + \ldots + \gamma_r \prod_{I_r} y_i$$

This expression may be rewritten as

$$\gamma_r + {}_rC_1\gamma_{r-1} + \ldots + {}_rC_j\gamma_{r-j} + \ldots + {}_rC_{r-1}\gamma_1 + {}_rC_r\gamma_0,$$

where γ_0 is defined as 0. Hence the general expression for α_r in terms of the γ's is

$$\alpha_r = \sum_{j=0}^{r} {}_rC_j\, \gamma_{r-j}, \quad r=0,\ldots,n. \qquad (2.2.3)$$

Similarly, the γ's may be written in terms of the α's, for example,

$$\gamma_1 = \alpha_1,$$

$$\gamma_2 = \alpha_2 - 2\alpha_1.$$

Claim The term γ_r in expression (2.2.2) may be written as a function of the α parameters defined in (2.2.1) in the following way

$$\gamma_r = \sum_{m=0}^{r} {}_rC_m \, (-1)^m \, \alpha_{r-m} \qquad (2.2.4)$$

where α_0 is defined as 0.

Proof: Expression (2.2.3) may be written as

$$\gamma_r = \alpha_r - \sum_{j=1}^{r} {}_rC_j \, \gamma_{r-j} \qquad (2.2.5)$$

By mathematical induction, assume that (2.2.4) holds for all values of γ_j such that j is less than r. Then

$$\gamma_{r-j} = \sum_{h=0}^{r-j} {}_{r-j}C_h \, (-1)^h \, \alpha_{r-j-h}, \quad j = 1,\ldots,r \qquad (2.2.6)$$

Substitute (2.2.6) into (2.2.5) to get:

$$\gamma_r = \alpha_r - \sum_{j=1}^{r} \sum_{h=0}^{r-j} {}_rC_j \, {}_{r-j}C_h \, (-1)^h \, \alpha_{r-j-h} \qquad (2.2.7).$$

Consider the coefficient of α_{r-m} in the right hand side of (2.2.7), namely the sum of coefficients of α_{r-j-h} where

$$j + h = m,$$

that is,

$$j = 1,\ldots,m$$

and

$$h = m-1,\ldots,0.$$

The required coefficient is

$$- \sum_{i=1}^{m} {}_rC_i \, {}_{r-i}C_{m-i} \, (-1)^{m-i}$$

$$= (-1)^{m+1} \sum_{i=1}^{m} \frac{r!}{i!(r-i)!} \, \frac{(r-i)!}{(m-i)!(r-m)!} \, (-1)^{-i}$$

$$= \frac{r!}{(r-m)!m!} (-1)^{m+1} \sum_{i=1}^{m} \frac{m!}{(m-i)!i!} (-1)^i$$

$$= {}_rC_m (-1)^{m+1} [\sum_{i=0}^{m} {}_mC_i (-1)^i - 1]. \qquad (2.2.8)$$

But

$$\sum_{i=0}^{m} {}_mC_i (-1)^i$$

is the binomial expansion of $(1 - 1)^m$ which is 0, so that the coefficient of α_{r-m} given by (2.2.8) is

$${}_rC_m (-1)^m,$$

and (2.2.7) becomes

$$\sum_{m=0}^{r} {}_rC_m (-1)^m \alpha_{r-m}.$$

The conditional distribution of Y defined in (2.2.1) allows one to examine the marginal distribution of each Y variable and, moreover, the joint distribution of any pair of Y's, say Y_1 and Y_2, and in particular the moments of the joint distribution. If one uses $P(i_1,\ldots,i_n)$ to denote the joint (conditional) distribution of (2.2.1) then

$$EY_1 = \sum_{i_2,\ldots,i_n} P(1,i_2,\ldots,i_n)$$

$$= \exp(\beta x_1) \{\exp(\alpha_1) + \exp(\alpha_2) \sum_{i>1} \exp(\beta x_i)$$

$$+ \exp(\alpha_3) \sum_{j>i>1}^{n} \exp[\beta(x_i + x_j)] + \ldots$$

$$+ \exp(\alpha_n) \exp(\beta \sum_{i>1}^{n} x_i)\} / d,$$

where d was defined in (2.2.1). One may then write EY_1 as

$$b_1 \sum_{h=0}^{n-1} (a_{h+1} c_h) / \sum_{h=0}^{n} (a_h d_h),$$

where

$$b_i = \exp(\beta x_i), \; i=1,\ldots,n$$

$$b_0 = 1,$$

$$a_i = \exp(\alpha_i), \; i=1,\ldots,n$$

$$a_0 = 1,$$

$$c_h = \sum_{i_h > i_{h-1} > \ldots > i_1 > 1} \exp[\beta(\sum_{j=i_1}^{i_h} x_j)]$$

and

$$d_h = \sum_{i_h > i_{h-1} > \ldots > i_1} \exp[\beta(\sum_{j=i_1}^{i_h} x_j)].$$

Hence $1 - EY_1$ may be written

$$[d - b_1 \sum_{h=0}^{n-1} (a_{h+1} c_h)]/d$$

$$= \sum_{h=0}^{n-1} (a_h c_h) / d$$

59

The variance of Y_1, denoted by VY_1, which, because of the binary nature of Y, can be written as $EY_1(1 - EY_1)$, is

$$= [b_1 \sum_{h=0}^{n-1} (a_{h+1} c_h) \sum_{h=0}^{n-1} (a_h c_h)] / d^2.$$

Similarly VY_2 can be written

$$= [b_1 \sum_{h=0}^{n-1} (a_{h+1} e_h) \sum_{h=0}^{n-1} (a_h e_h)] / d^2.$$

where

$$e_h = \sum_{i_h > i_{h-1} > \ldots > i_1 \neq 2} \exp(\beta \sum_{j=i_1}^{i_h} x_j)$$

This notation also simplifies the writing of the moments of the joint distribution, for example,

$$EY_1Y_2 = b_1 b_2 \{\alpha_2 \sum_{i>2} \exp(\beta x_i) + \alpha_3 \sum_{j>i>2} \exp[\beta(x_i + x_j)]$$

$$+ \ldots + \alpha_n \exp[\beta(\sum_{i>2} x_i)]\} / d$$

and $\mathrm{cov}(Y_1, Y_2)$

$$= b_1 b_2 \{d \sum_{h=0}^{n-2} (a_{h+2} f_h) - [\sum_{h=0}^{n-1} (a_{h+1} c_h) \sum_{h=0}^{n-1} (a_{h+1} e_h)]\} / d^2, \quad (2.2.9)$$

where

$$f_h = \sum_{i_h > i_{h-1} > \ldots > i_1 > 2} \exp[\beta(\sum_{j=i_1}^{i_h} x_j]$$

We are interested in the expression in parenthesis in (2.2.9), which can be written

$$[\sum_{h=0}^{n} (a_h d_h) \sum_{h=0}^{n-2} (a_{h+2} f_h)] - [\sum_{h=0}^{n-1} (a_{h+1} c_h) \sum_{h=0}^{n-1} (a_{h+1} e_h)]. \quad (2.2.10)$$

This expression may be rewritten in terms of the powers of $\exp(\beta x_i)$, in particular, in terms of the coefficients of

$$f_i f_j,\ i,j=1,\ldots,n.$$

First we write c_h, d_h and e_h in terms of f_h, f_{h-1} and f_{h-2}, that is,

$$c_h = f_h + b_2 f_{h-1},$$

$$e_h = f_h + b_1 f_{h-1},$$

$$d_h = f_h + (b_1+b_2)f_{h-1} + b_1 b_2 f_{h-2}.$$

These equations can be substituted into (2.2.10) to produce

$$\sum_{h=0}^{n} a_h(f_h + b_1 f_{h-1} + b_2 f_{h-1} + b_1 b_2 f_{h-2}) \sum_{h=0}^{n-2} a_{h+2} f_h$$

$$- \sum_{h=0}^{n-1} a_{h+1}(f_h + b_2 f_{h-1}) \sum_{h=0}^{n-1} a_{h+1}(f_h + b_1 f_{h-1}).$$

The coefficient of $(f_i f_j + f_j f_i)$, except when

$$i=j,$$

can be written

$$\begin{aligned}
&(a_i + a_{i+1}b_1 + a_{i+1}b_2 + a_{i+2}b_1 b_2)a_{j+2} \\
&\quad + (a_j + a_{j+1}b_1 + a_{j+1}b_2 + a_{j+2}b_1 b_2)a_{i+1} \\
&\quad - (a_{i+1} + a_{i+2}b_2)(a_{j+1} + a_{j+2}b_1) \\
&\quad - (a_{j+1} + a_{j+2}b_2)(a_{i+1} + a_{i+2}b_1) \\
&= a_i a_{j+2} + a_j a_{i+2} + (a_{i+1}a_{j+2} + a_{j+1}a_{i+2})b_1 \\
&\quad + (a_{i+1}a_{j+2} + a_{j+1}a_{i+2})b_2 + 2a_{i+2}a_{j+2}b_1 b_2 \\
&\quad - [2a_{i+1}a_{j+1} + (a_{i+1}a_{j+2} + a_{j+1}a_{i+2})b_1 \\
&\quad + (a_{i+1}a_{j+2} + a_{j+1}a_{i+2})b_2 + 2a_{i+2}a_{j+2}b_1 b_2] \\
&= a_i a_{j+2} + a_j a_{i+2} - 2a_{i+1}a_{j+1}.
\end{aligned}$$

Hence, one may say that, for

$$i \neq j,$$

the coefficient of $f_i f_j$ is

$$a_i a_{j+2} - a_{i+1} a_{j+1}. \qquad (2.2.11)$$

Moreover, when

$$i=j$$

the coefficient of $f_i f_j$ is

$$(a_i + a_{i+1}b_1 + a_{i+1}b_2 + a_{i+2}b_1 b_2)a_{i+2}$$

$$- (a_{i+1} + a_{i+2}b_2)\ (a_{i+1} + a_{i+2}b_2)$$

$$= a_i a_{i+2} - a_{i+1}^2,$$

which is merely (2.2.11) with j set equal to i. This expression (2.2.11) holds for all values of i and j, in particular for

$$i, j > n-3$$

provided one defines

$$a_{n+1} = a_{n+2} = 0.$$

Note that for

$$i = j + 1$$

the coefficient of $f_i f_j$ in the expansion of $cov(Y_1, Y_2)$ is zero.

2.3 Rosner's correlated logistic model

Rosner(1983,1984) proposed a correlated logistic model which can be interpreted as the result of imposing constraints on the parameters $\{a_j, j=1,\ldots,n\}$ of the general correlated logistic model (2.2.1). In particular, he suggested that these n parameters be reduced to 2 parameters, a and b, by means of the expressions

$$\alpha_j = (a)_j/(b+n-j)_j \quad ,j=1 \ldots,n$$

where

$$(c)_j = c(c+1)\ldots(c+j-1).$$

This set of constraints does not simplify the expression for the model according to Dempster's formulation (2.2.2), nor the expression of γ_r in terms of the α's (2.2.4). However, in the covariance of Y_1 and Y_2, we may write the expression for the coefficient of $f_i f_j$ (2.2.11) as

$$a_i a_{j+2} - a_{i+1} a_{j+1}$$

$$= \frac{(a)_i}{(b+n-i)_i} \frac{(a)_{j+2}}{(b+n-j-2)_{j+2}} - \frac{(a)_{i+1}}{(b+n-i-1)_{i+1}} \frac{(a)_{j+1}}{(b+n-j-1)_{j+1}}$$

$$= \frac{(a)_i \, (a)_{j+1}}{(b+n-i)_i (b+n-j-1)_{j+1}} \left[\frac{(a+j+1)}{(b+n-j-2)} - \frac{(a+i)}{(b+n-i-1)} \right]$$

$$= \frac{(a)_i (a)_{j+1}}{(b+n-i-1)_{i+1} (b+n-j-2)_{j+2}} [(a+b+n-1)(j-i+1)]. \quad (2.3.2)$$

Rosner's model is of considerable interest because at certain values of the covariates, it reduces to some familiar models, in particular, a generalised hypergeomtric and the beta-binomial. For example, when the covariates are cluster-specific, that is,

$$\mathbf{x}_1 = \mathbf{x}_2 = \ldots = \mathbf{x}_n = \mathbf{x},$$

then expression (2.2.1) becomes

$$\log \Pr(\mathbf{Y}=\mathbf{y}/\mathbf{x}) = \alpha_r + r\boldsymbol{\beta}'\mathbf{x} - \log d,$$

that is,

$$\Pr(\mathbf{Y}=\mathbf{y}) = \exp(\alpha_r + r\gamma) / \sum_{j=0}^{n} {}_nC_j \exp(\alpha_j + j\gamma),$$

where

$$\gamma = \boldsymbol{\beta}'\mathbf{x}.$$

If R is a random variable that denotes the number of y's that are 1, that is,

$$R = \sum_{i=1}^{n} Y_i$$

then the probability mass function of R is

$$\Pr(R=r) = {}_nC_r \exp(\alpha_r + r\gamma) / \sum_{j=0}^{n} {}_nC_j \exp(\alpha_j + j\gamma). \qquad (2.3.3)$$

Using Rosner's constraints (2.3.1), we get

$$\exp(\alpha_r + r\gamma) = \exp(r\gamma)\,(a)_r/(b+n-r)_r$$

so that (2.3.3) becomes

$$\Pr(R=r) = {}_nC_r \exp(r\gamma) \frac{(a)_r/(b+n-r)_r}{\sum {}_nC_j (a)_j \exp(j\gamma)/(b+n-j)_j} \qquad (2.3.4)$$

Since

$$(b)_n/(b+n-j)_j = (b)_{n-j},$$

multiplying the numerator and denominator of (2.3.4) by $(b)_n$ yields

$$\Pr(R=r) = {}_nC_r \frac{(a)_r (b)_{n-r}\, g^r}{\sum {}_nC_j (a)_j (b)_{n-j}\, g^j}, \qquad (2.3.5)$$

where

$$g = \exp(\gamma).$$

Expression (2.3.5) is the probability mass function of a generalised hypergeometric distribution similar to the 'extended' hypergeometric distribution described by Johnson and Kotz (1969, p.160-2).

When either **x** or **β** is the zero vector, we have

$$\gamma = 0$$

and

$$g = 1,$$

so that the right hand side of (2.3.5) becomes

$$_nC_r\ (a)_r (b)_{n-r}\ /\ \Sigma\ _nC_j\ (a)_j (b)_{n-j} \qquad (2.3.6).$$

Consider the denominator of (2.3.6).

Claim:
$$\sum_{j=0}^{n} {}_nC_j\ (a)_j (b)_{n-j} = (a+b)_n \qquad (2.3.7)$$

Proof: the left hand side of (2.3.7) may be written

$$\sum_{j=0}^{n} \frac{n!}{j!(n-j)!} (a)_j\ (b)_{n-j}$$

$$= n!\ \Sigma\ \frac{(a)_j}{j!} \frac{(b)_{n-j}}{(n-j)!}$$

$$= n!\ (a+b+n-1)_n$$

(see, for example, Feller(1968, p65))

$$= (a+b)_n.$$

Hence the probability mass function of R (2.3.6) may be written

$$_nC_r\ (a)_r (b)_{n-r}/(a+b)_n.$$

This is the probability mass function of the beta-binomial distribution (see Section 1.4.2). Hence, when the covariates are cluster-specific and equal to the zero vector, or when covariates have no effect (that is, the regression coefficients are all zero) the distribution of R is beta-binomial.

It is now possible to clarify the relationship between Rosner's model and that proposed for cluster-specific covariates by Williams(1982, see also section 1.5.2). Williams' model is always a beta-binomial because the cluster-specific covariates are part of a parameter of the beta-binomial distribution. If we write μ as the mean of the random varible R, then

$$\mu = g/(1+g),$$

where

$$g = \exp(\alpha+\beta'\mathbf{x}).$$

Thus, when the covariate or the regression coefficient is zero, R is a beta-binomial with mean

$$\mu = \exp(\alpha)/[1+\exp(\alpha)]$$

and correlation ρ. Similarly, for Rosner's model in this case, there are also two paremeters, a and b, or equivalently, α_1 and α_2, so that

$$\mu = a/(a+b)$$

and

$$\rho = 1/(a+b+1).$$

Thus α of Williams' model may be written as

$$\alpha = \log(a/b).$$

The parameters of Williams' model may be also written in terms of the parameters, α_1 and α_2, but the expressions are more complex. Thus, when the covariate or the regression parameter is zero, the two models are equivalent. However, in general, when the covariates are cluster-specific, Rosner's and Williams' models are not equivalent.

Next consider the effect of assuming cluster-specific covariates on the moments, in particular, the variances and covariances of Y_1 and Y_2. For cluster-specific covariates,

$$VY_1 = VY_2 = VY$$

$$= g \sum_{h=0}^{n-1} (a_{h+1} c_h) \sum_{h=0}^{n-1} (a_h c_h) / d^2 .$$

where

$$g = \exp(\boldsymbol{\beta}' \mathbf{x})$$

$$c_h = {}_{n-1}C_h \, g^h$$

so that, under Rosner's model,

$$\sum a_{h+1} c_h = \sum (a)_{h+1} \, {}_{n-1}C_h \, g^h / (b+n-h-1)_{h+1}$$

$$= \frac{(n-1)!\, a}{(b)_n} \sum_{h=0}^{n-1} \binom{a+h}{h} \binom{b+n-h-2}{n-h-1} g^h .$$

Similarly

$$\sum a_h c_h = \sum (a)_h \, {}_{n-1}C_h \, g^h / (b+n-h)_h$$

$$= \frac{(n-1)!}{(b+1)_{n-1}} \sum_{h=0}^{n-1} \binom{a+h-1}{h} \binom{b+n-h-1}{n-h-1} g^h .$$

and

$$d = \sum (a)_h \, {}_nC_h \, g^h / (b+n-h)_h$$

$$= \frac{n!}{(b)_n} \sum_{h=0}^{n-1} \binom{a+h-1}{h} \binom{b+n-h-1}{n-h-1} g^h .$$

The covariance between Y_1 and Y_2 may be evaluated by using (2.2.9) with coefficients of $f_i f_j$ given by (2.3.2), as well as the fact that

$$f_i = {}_{n-2}C_i \, g^i .$$

Hence we have

$$\mathrm{Cov}(Y_1,Y_2) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \left[\frac{(a)_i (a)_{j+1}}{(b+n-i-1)_{i+1}(b+n-j-2)_{j+2}} (a+b+n-1)(j-i+1) \times {}_{n-2}C_i \, {}_{n-2}C_i \, g^{i+j+2}\right]/d^2 \quad (2.3.8)$$

where d was defined above.

Now when the cluster-specific covariate is null or the regression coefficient is zero, then it has been shown that R, the sum of the Y's, is a beta-binomial covariate. This implies that

$$\mathrm{corr}(Y_1,Y_2) = 1/(a+b+1)$$

This result can be verified by substituting in the expressions above for VY and $\mathrm{cov}(Y_1,Y_2)$ and calculating

$$\mathrm{corr}(Y_1,Y_2) = \mathrm{cov}(Y_1,Y_2)/VY.$$

It should be noted, however, that, in the usual case, that is, when the covariates are not null and/or the regression coefficients are non-zero, then the correlation between Y_1 and Y_2, under the correlated logistic model, and also under Rosner's simpler model, is not a constant, but is a complicated function of the model parameters (α and β), as well as of the values of the covariates themselves.

2.4 The conditional model

When the correlated logistic model holds for Y, α may be considered as a nuisance parameter in inferences about β. One method for removing α from the inference problem is to construct a conditional likelihood function, that is, to find a statistic, and to condition on it, so as to remove the nuisance parameters from the likelihood function (see, for example, Cox and Hinkley(1974)). In dealing with the correlated logistic, we follow the method of Breslow, Day et al.(1978) (who were working with a stratified logistic model (see section 1.7.3)), and, for each cluster, condition on the number of occurrences (that is, the number of y's that have value 1). Assume that there are r occurrences so that

$$\Pr(\mathbf{Y}=\mathbf{y}/X, r \ y\text{'s are } 1)$$

$$= \frac{\exp[\alpha_r + \Sigma (y_i \boldsymbol{\beta}'\mathbf{x}_i)]}{\sum_{j=1}^{{}_nC_r} [\exp(\alpha_j + \sum_{i\epsilon I_j} \boldsymbol{\beta}'\mathbf{x}_i)]}$$

$$= \exp(\ \Sigma\ y_i\ \boldsymbol{\beta}'\mathbf{x}_i)/\ \ \Sigma\ \exp(\ \Sigma\ \boldsymbol{\beta}'\mathbf{x}_i)$$

If we relabel the units such that the first r have the y value of 1, then we have the probability

$$\exp(\sum_{i=1}^{r} \boldsymbol{\beta}'\mathbf{x}_i)/\sum_{j=1}^{{}_nC_r} \exp(\sum_{i\epsilon I_j} \boldsymbol{\beta}'\mathbf{x}_i). \qquad (2.4.1)$$

This is the same conditional probability obtained by Breslow and Day (1980, p202-5) for the stratified logistic model.

When the clusters are all of size 2 (and we ignore clusters with both y values the same, as will argued in Chapter 3), expression (2.4.1) becomes

$$\Pr(Y_1=1, Y_2=0) = \exp(\boldsymbol{\beta}'\mathbf{x}_1)/[\exp(\boldsymbol{\beta}'\mathbf{x}_1) + \exp(\boldsymbol{\beta}'\mathbf{x}_2)],$$

which is exactly the simple Breslow-Day conditional model(Breslow, Day et al.(1978, p.302)).

CHAPTER 3

Asymptotic behaviour of estimators

3.1 Introduction

In this chapter, we consider the asymptotic properties of various estimators of the regression coefficients of the correlated logistic model, either as the full model or as Rosner's simplification of the correlated logistic. In section 3.2 we describe the estimators of interest, including those obtained from maximum likelihood, conditional maximum likelihood, and standard logistic regression with and without the incorporation of dummy variables to account for the clustering. In section 3.3 we establish the conditions under which the estimators are consistent (asymptotically unbiased), showing that the use of standard logistic regression, with or without dummy variables, is in general asymptotically biased. In section 3.4 we derive the general formulae for the asymptotic variance of these estimators and investigate further for the case

$$\beta = 0.$$

In this special case we derive the misspecification factor for the usual and dummy variables estimators.

3.2 The estimators

The estimators to be considered in this thesis are:

1. the maximum likelihood estimator, denoted by $\hat{\beta}_m$, or when the constraints specified by Rosner are imposed, Rosner's maximum likelihood estimator, denoted by $\hat{\beta}_R$.

2. the conditional maximum likelihood estimator, constructed in the manner proposed by Breslow, Day et al.(1978), in the context of case-control studies. In section 2.4 we obtained, for the correlated logistic regression model in a single cluster, the same likelihood function that Breslow and Day produced by conditioning in the stratified logistic model. We propose to minimize this likelihood function over all k clusters. This estimator so obtained is denoted by $\hat{\beta}_c$.

3. the estimator under the usual (or uncorrelated) logistic model, denoted by $\hat{\beta}_u$.

4. the estimator under the dummy variables model, whereby indicator variables are added to the usual logistic model to account for clustering, in particular, in the i'th cluster, the dummy (or indicator) variable V_i, takes on the value 1, whereas in all other clusters, V_i takes on the value 0. This logistic dummy varibles model has been studied in educational testing (Andersen(1973),

p66-9), econometrics (Chamberlain(1980)) and medicine (Breslow, Day et al(1978), Smith and Pike(1980), etc.), where in case-control studies it is as known as stratified analysis. The dummy variables estimator is denoted by $\hat{\beta}_d$.

Another approach that could be considered is that developed by Williams(1982) in which the error distribution is assumed to be betabinomial (see section 1.5). As was shown in section 2.3, Willams' model is the same as Rosner's only when the covariate or the regression parameter is zero; otherwise they are different models. However, Williams', model only allows cluster-specific covariates. In this thesis we are interested in models for mixed cluster- and unit-specific covariates, or at the least, in models for unit-specific covariates; hence Williams' model will not be considered further.

3.3 Consistency

3.3.1 Maximum likelihood estimator

As stated in section 1.8.2, we follow Hauck(1982) and make the assumption that the joint distribution of the observations on one cluster can be written

$$f(y,x;\theta_1,\theta_2) = f(y|x;\theta_1)\, g(x;\theta_2) \qquad (3.3.1.1)$$

where θ_1 and θ_2 are the parameters of the distribution and

$$\theta_1 = (\alpha,\beta),$$

that is , the joint distribuion of the dependent variables and the covariates can be written as two functions, the conditional distribution of the dependent variable (usually modelled by the correlated logistic model), and the marginal distribution of the covariates; moreover, it is assumed that the distribution of the covariates does not depend on the parameters of the conditional distribution of the dependent variable. If we further assume that the joint distribution (3.3.1.1) has the usual regularity properties (see, for example, Cox and Hinkley(1974), p281), and that the first three moments of X exist and are finite (see section 1.8.2), then it follows that $\hat{\beta}_m$ (or $\hat{\beta}_R$, if the Rosner constraints have been imposed) is a consistent(asymptotically unbiased) and asymptotically normal estimator of β, the regression parameter.

3.3.2 The conditional maximum likelihood estimator

As stated in section 1.8.3, Andersen(1970, 1973) studied the properties of the conditional maximum likelihood estimator. Let us consider his arguments as applied to the correlated logistic estimator. First we assume that the covariates are fixed (or, equivalently, that we may condition on their observed values).

Next we condition on a sufficient statistic. In section 2.3 a conditional distribution was obtained, (following Breslow, Day et al.(1978)) by conditioning on R, the number of occurrences (y values which are equal to 1). Since the distribution conditional on R is independent of α, R is a sufficient statistic for α. For a single cluster, R is a minimal sufficient statistic. However, for a set of k clusters of size n, the minimal sufficient statistic for α is the set $\{F_1,\ldots,F_n\}$, where F_i is the number of clusters with i occurrences. Nevertheless, a conditional estimator, namely $\hat{\beta}_c$, may be produced by minimizing the conditional likelihood of Breslow and Day over all k clusters. The results of Andersen(1970) on the asymptotic distribution of condtional maximum likelihood estimators do not require the use of a minimal sufficient statistic. However, when $\hat{\beta}_c$ is used, information about β is lost; this loss of information will be discussed later in this Chapter, and also in Chapter 4.

The conditional likelihood of Breslow and Day conforms to the conditioning required by Andersen(1970), that is, $f(y|x;\theta_1)$, as defined in (3.3.1.1), can be written as

$$f(y|x;\theta_1) = f(y|x,r;\beta)\ h(r|x;\alpha,\beta)$$

where h(.) is the distribution of R and $f(y|x,r;\beta)$ is the conditional distribution obtained in section 2.4. If we assume the same partitioning of the distributon of x and y that occurs in (3.3.1.1), then we have the following conditional distributon function

$$f(y,x|r;\beta,\theta_2) = f(y|x,r;\beta)\ g(x;\theta_2)$$

If we assume that the function $f(y,x|r;\beta,\theta_2)$ satisfies the regularity conditions of Andersen, then $\hat{\beta}_c$ is a consistent (asymptotically unbiased) and asymptotically normal estimator of β, the regression parameter.

74

3.3.3 The usual estimator

The usual estimator is that estimator obtained by ignoring the correlation structure within each cluster and constructing estimating equations for the usual logistic model. The estimator is denoted by $\hat{\beta}_u$. The method employed in studying the consistency of $\hat{\beta}_u$ was suggested by a lemma of Brillinger(1975, p94-6). A similar approach was used on a problem in applied probability by Kulperger(1979). In general, one examines the asymptotic behaviour of the estimating equations, in this case, the incorrect maximum likelihood equations, under the correct model.

In particular, the following approach is employed:

1. let $l(\theta)$ denote the log likelihood for a single cluster, and let

$$l_i(\theta),\ i=1,\ldots,k,$$

   be the log likelihood functions for the k clusters in a sample, and let

$$L_k(\theta) = \frac{1}{k}\sum_{i=1}^{k} l_i(\theta)$$

   denote the average log likelihood of the sample.

2. under regularity conditions described in section 1.8.4, a law of large numbers holds so that

$$L_k(\theta) \xrightarrow{p} L(\theta) = E(l(\theta)),$$

where the expectation is with respect to the joint distribution of Y, the dependent variable, and X, the covariate; in this thesis Y has the correlated logistic distribution. Hence the value of θ which minimizes $L(\theta)$ is the limiting value of the sequence of estimators, $\{\theta_k,\ k=1,\ldots\}$, which minimize the functions $\{L_k,\ k=1,\ldots\}$. Denote the limiting value of the sequence $\{\theta_k\}$ by θ_0.

3. to find the limiting value, θ_0, of the sequence of estimators, take derivatives of $L(\theta)$ with respect to θ, yielding, $S(\theta)$, the limiting form of the estimating equations under the correct model. The solution to the equation

$$S(\theta) = 0$$

yields the value of θ_0.

4. investigate the values of θ_0 for various values of the true parameter θ.

For the usual model, the log likelihood for a single unit is written

$$l_j(\alpha,\beta) = \alpha y_j + \beta x_j y_j - \log(d_j),$$

where

$$d_j = 1 + \exp(\alpha + \beta x_j).$$

Note that, although $\boldsymbol{\beta}$ and $\mathbf{x}_j$ are both vectors, because the veetor attributes are not used at this time, we use the terms β and x_j, without indication of their vector properties.

The log likelihood for a complete cluster may be written

$$l(\alpha,\beta) = \sum_{j=1}^{n} l_j(\alpha,\beta)$$

$$= \alpha \sum_{j=1}^{n} y_j + \beta \sum_{j=1}^{n} x_j y_j - \log(d),$$

where

$$d = \prod_{j=1}^{n} d_j .$$

The average (over k) log likelihood for k clusters may be written

$$L_k(t) = \frac{1}{k} \sum_{i=1}^{k} l_i(\alpha,\beta)$$

$$= [\alpha \Sigma y_{ij} + \beta \Sigma x_{ij} y_{ij} - \log(d_i)] /k.$$

Under the regularity conditons described in section 1.8.4, $L_k(t)$ has the following limit in probability

$$\alpha \sum_{i=1}^{k} E(Y_i) + \beta \sum_{i=1}^{k} E(X_i Y_i) - E(\log D),$$

which, since the marginal distributions of the (Y_i, X_i) are the same for each i=1,...,k, can be written

$$k\alpha EY + k\beta EXY - E(\log D),$$

where the expectation is with respect to joint distribution of X and Y. However, EY and EXY may be simplified in the following way

$$EY = E_X E_{Y|X}(Y)$$

$$= E_X F,$$

where F is the probability that Y is 1, conditional on value of X, that is,

$$F = \Pr(Y=1|X).$$

Note that F is a function of all values of the covariate in a cluster. Similarly it may be shown that

$$EXY = E_X(XF)$$

Thus the limiting form of the log likelihood may be written as

$$L(\alpha,\beta) = k\alpha E_X F + k\beta E_X(XF) - E(\log D).$$

Taking derivatives with respect to α we have

$$\frac{\partial L}{\partial \alpha} = kE_X F - \frac{\partial}{\partial \alpha} E(\log D),$$

Assuming the interchangeability of differentiation and integration, we get

$$\frac{\partial}{\partial \alpha} E(\log D) = E\left(\frac{\partial D}{\partial \alpha} / D \right).$$

Since

$$D = \prod_{i=1}^{n} [1 + \exp(\alpha + \beta X_i)],$$

$\partial D/\partial\alpha$ may be written

$$\exp(\alpha+\beta X_1) \prod_{i\neq 1} [1+\exp(\alpha+\beta X_i)]$$

$$+ \exp(\alpha+\beta X_2) \prod_{i\neq 2} [1+\exp(\alpha+bX_i)] + \ldots$$

so that

$$\left(\frac{\partial D}{\partial \alpha}\right) / D = \frac{\exp(\alpha+\beta X_1)}{1+\exp(\alpha+\beta X_1)} + \frac{\exp(\alpha+bX_2)}{1+\exp(\alpha+\beta X_2)} + \ldots$$

$$= \sum_{i=1}^{n} P_i$$

where

$$P_i = \exp(\alpha+\beta X_i)/[1+\exp(\alpha+\beta X_i)],$$

that is, P_i is the probability that Y is 1 under the usual logistic model. Hence we may write $\partial L/\partial\alpha$ as

$$nE_X F - nE_X P \qquad (3.3.3.1)$$

Next we calculate the derivatives with respect to $\boldsymbol{\beta}$, that is,

$$\partial L/\partial\beta_j,\ j=1,\dots p.$$

Now

$$\frac{\partial L}{\partial\beta_j} = n\,E_X(X_j F) - \frac{\partial}{\partial\beta_j}E(\log D).$$

Again, assuming interchangability of differentiation and integration, we get

$$\frac{\partial}{\partial\beta_j}E(\log D) = E\left(\frac{\partial D}{\partial\beta_j}/D\right)$$

$$= E(X_{1j}\exp(\alpha+\boldsymbol{\beta}'\mathbf{X}_1)\prod_{i\neq 1}\exp(\alpha+\boldsymbol{\beta}'\mathbf{X}_i) + \dots /D),$$

where the vector notation has been re-introduced for clarity. This expression further reduces to

$$E\left(\sum_{i=1}^{n} X_{ij}P_i\right) = n\,E(X_j P).$$

Hence $\partial L/\partial\boldsymbol{\beta}$ may be written

$$nE_X(\mathbf{X}F) - nE_X(\mathbf{X}P) \qquad (3.3.3.2)$$

where $\mathbf{X}$ denotes the $(1\times p)$ random vector for a single unit and X is the random matrix for a cluster of size n, that is,

$$X = (\mathbf{X}_1,\dots,\mathbf{X}_n).$$

The limiting form of the estimating equations is given by setting each

The difference between this simple version of the correlated logistic model and the simple version of Cox's multivariate logisitic model(1.3.5) can be summarised as follows

1. the correlated logistic allows for two values of the covariates, x_1 and x_2, to be associated with each set of values of the dependent variables, (y_1, y_2),

2. the assumption

$$x_1 = x_2 = x$$

   changes the correlated logistic into a version of the multivariate logistic,

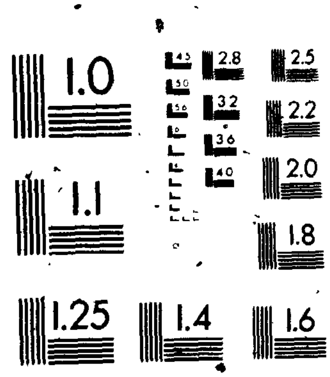
3. the correlated logistic may have some covariates measured at the same level as each y value, that is at the unit level, but may also have some covariates measured at the same level as the pair (y_1, y_2), that is, at the cluster level.

The correlated logistic is proposed for the following types of data

1. data taken on individuals in a sample of families
2. data taken on eyes, ears, hands, etc., in a sample of individuals
3. data arising from repeated measures on the same subject

2

MICROCOPY RESOLUTION TEST CHART
NBS 1010a
ANSI and ISO TEST CHART No 2



equation in (3.3.3.1) and (3.3.3.2) equal to zero, that is,

$$E_X(F) = E_X(P)$$

and

$$E_X(\mathbf{X}F) = E_X(\mathbf{X}P) \qquad (3.3.3.3)$$

where F is a function of the parameters of the correct model (the correlated logistic) and P is a function of the parameters of the incorrect model (in this case, the usual logistic).

Equation (3.3.3.3) is studied in some detail in Chapter 4 to determine the behaviour of estimators for simple distributions of X. In this section, we consider the results when the true regression coefficient is zero, that is,

$$\boldsymbol{\beta} = \mathbf{0}.$$

In this case, F is a constant with respect to X and equation (3.3.3.3) simplifies to

$$F = E_X(P)$$

$$F\, E_X(\mathbf{X}) = E_X(\mathbf{X}P) \qquad (3.3.3.4)$$

Substituting the first equation of (3.3.3.4) into the remainder, we get

$$E_X(\mathbf{X})\, E_X(P) = E_X(\mathbf{X}P).$$

By repeated application of the integral form of Holder's inequality (see, for example, Rao(1973), p75), we can show that this implies that P is a constant with respect to $\mathbf{X}$, that is, $\boldsymbol{\beta}_0$, the limiting value of $\hat{\boldsymbol{\beta}}_u$, is $\mathbf{0}$; in other words,

$$\hat{\beta}_u \xrightarrow{P} 0;$$

hence, in this case, $\hat{\beta}_u$ is a consistent estimator of the parameter β.

In this limiting case, P is not a function of X, and the first equation of (3.3.3.4) implies that

$$P = F,$$

that is,

$$\exp(\alpha_0)/[1+\exp(\alpha_0)] = F,$$

where F is a function of the parameter α of the correlated logistic model. We may write

$$\alpha_0 = \text{logit}(F).$$

Another condition under which $\hat{\beta}_u$ is a consistent estimator of β occurs when the usual logistic model is correct and there in no correlation between the Y's within a cluster. In this case, the parameters of the correlated model may be written in term of a single α parameter, namely,

$$\alpha_r = r\alpha, \ r=1,\ldots,n.$$

It is conjectured that, in this case, estimation of β using the correlated logistic model will suffer a loss of efficiency due to the estimation of the n parameters $\{\alpha_r, \ r=1,\ldots,n\}$ instead of the single parameter α.

The results given in this secton hold for finite n (cluster size). In chapter 4, some numerical results are given for the sign and bias of $\hat{\beta}_u$ as n becomes large.

The remaining question considered in this secton is whether $\hat{\beta}_u$ can estimate consistently a subset of the elements of the parameter β when the remainder of the β's are non-zero. We now show some very restrictive conditions when this is possible. First we write β as two subvectors β_a and β_b, namely,

$$\beta = \begin{pmatrix} \beta_a \\ \beta_b \end{pmatrix}$$

and show that, when β_a is zero, it is possible under certain conditions to estimate the subvector β_a consistently. We write equation(3.3.3.3) as

$$\begin{aligned} E_X(F) &= E_X(P) \\ E_X(X_a F) &= E_X(X_a P) \qquad (3.3.3.5) \\ E_X(X_b F) &= E_X(X_b P) \end{aligned}$$

where X has been partitioned in a manner similar to β, that is,

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix}.$$

Let us assume that β_a is zero, so that F is not a function of X_a. The first equation of (3.3.3.5) becomes

$$E_{X_b}(F) = E_X(P) \qquad (3.3.3.6)$$

where the matrix X has been partitioned in a manner similar to β and the vector X, namely,

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix}.$$

The second equation in (3.3.3.5) can be written

$$E_{X_a}(\mathbf{X}_a)\, E_{X_b|X_a}(F) = E_X(\mathbf{X}_a P).$$

If X_a is distributed independently of X_b, this becomes

$$E_{X_a}(\mathbf{X}_a)\, E_{X_b}(F) = E_X(\mathbf{X}_a P).$$

Substituting from (3.3.3.6) into this equation, we get

$$E_{X_a}(\mathbf{X}_a)\, E_X(P) = E_X(\mathbf{X}_a P),$$

but, since X_a is a subset of X, this may be written

$$E_X(\mathbf{X}_a)\, E_X(P) = E_X(\mathbf{X}_a P),$$

which, from Holder's inequality (see Rao(1975), p75), implies that P is not a functon of $\mathbf{X}_a$, that is,

$$\hat{\boldsymbol{\beta}}_{u_a} \xrightarrow{P} 0,$$

In summary, if

$$\boldsymbol{\beta}_a = \mathbf{0},$$

and, if X_a is distributed independently if X_b, then $\hat{\boldsymbol{\beta}}_u$ consistently estimates the subvector $\boldsymbol{\beta}_a$. However, there is no assurance about consistency in estimation of the subvector $\boldsymbol{\beta}_b$. A numerical example in Chapter 4 suggests that $\boldsymbol{\beta}_b$ is only estimated consistently when

1. $\boldsymbol{\beta}_b = \mathbf{0}$,

   or

2. there is no residual correlation between the Y's, and the usual logistic model is correct.

The numerical example in Chapter 4 gives the sign and size of the bias of the components of the usual estimator (that is, of $\hat{\boldsymbol{\beta}}_{u_a}$ and $\hat{\boldsymbol{\beta}}_{u_b}$) when $\boldsymbol{\beta}_a$ is zero, in the case of a simple distribution for $\mathbf{X}$.

3.3.4 Dummy variables estimator

For the usual logistic model used with dummy variables to incorporate the clustering, the log likelihood of the j´th unit in the i´th cluster may be written

$$l_{ij}(\boldsymbol{\beta},\boldsymbol{\gamma}) = (\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{v}_i)z_{ij} - \log d_{ij}, \quad i=1,\dots,k,\ j=1,\dots,n, \qquad (3.3.4.1)$$

where $\boldsymbol{\beta}$ is the (p×1) vector of regression parameters,
x_{ij} is the vector of covariates for the j´th unit in the i´th cluster,
$\boldsymbol{\gamma}$ is the (k×1) vector of dummy variables parameters,
$\mathbf{v}_i$ is the (k×1) vector of dummy(indicator) variables such that

$$v_{ih} = \delta_{ih},$$

where δ_{ih} is the Kronecker delta,
d_{ij} is the normalising constant,
and z_{ij} is the observed value of the dependent variable.
As shown in Appendix B.1, clusters having all Y values the same cannot be used with the dummy variables approach. Thus the likelihood (3.3.4.1) is conditonal upon the fact that, in each cluster, at least one of the responses differs from the others. We define the outcome variable Z to reflect this situation. The properties of Z are further studied in Appendix B.2.

Because the number of parameters, in particular, the number of elements in the vector $\boldsymbol{\gamma}$, approaches infinity as k does, the approach used in the previous section (3.3.3) with the usual estimator is not used here, that is, we do not consider the limit of the average likelihood function as the number of clusters goes to infinity. In this section we deal with the indicator parameters before letting the number of clusters become large, and then consider the limit of the average estimating equation.

The log likelihood of the entire sample is

$$\sum_{i=1}^{k}\sum_{j=1}^{n} l_{ij}(\boldsymbol{\beta},\boldsymbol{\gamma})$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n} [\boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{v}_i)z_{ij} - \log(d_{ij})].$$

Taking derivatives with respect to β_h we get

$$\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ijh}\,(z_{ij} - p_{ij}),\ h=1,\dots p \qquad (3.3.4.2),$$

where

$$p_{ij} = \Pr(Z_{ij}=1|\mathbf{x}_{ij},\mathbf{v}_i).$$

Taking derivatives with respect to γ_h we get

$$\sum_{i=1}^{n}\sum_{j=1}^{n} v_{ih}\,(z_{ij} - p_{ij}),\ h=1\dots k,$$

$$= \sum_{i=1}^{k} v_{ih} \sum_{j=1}^{n} (z_{ij} - p_{ij}),\quad h=1,\dots,k.$$

However, by the definition of v, for each h, v_{ih} is non-zero only for i=h, that is, only for the h'th cluster so that this expression becomes

$$\sum_{j=1}^{n} (z_{ij} - p_{ij}), \quad i=1,\ldots k. \qquad (3.3.4.3)$$

The estimators $\hat{\beta}_d$ and $\hat{\gamma}_d$ are obtained by setting the derivatives (3.3.4.2) and (3.3.4.3) equal to zero, obtaining the equations

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \mathbf{x}_{ij} z_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{n} \mathbf{x}_{ij} p_{ij} \qquad (3.3.4.4)$$

$$\sum_{j=1}^{n} z_{ij} = \sum_{j=1}^{n} p_{ij}, \quad i=1 \ldots k \qquad (3.3.4.5)$$

where p_{ij} is the estimated probability that Z_{ij} is 1 under the dummy variables model. We may sum the second set of equations (3.3.4.5) to yield a single equation, namely,

$$\sum_{i=1}^{k} \sum_{j=1}^{n} z_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{n} p_{ij}. \qquad (3.3.4.6)$$

Next consider the average of each equation as k goes to infinity. Equation (3.3.4.4) becomes

$$nE(\mathbf{X}Z) = nE(\mathbf{X}P) \qquad (3.3.4.7)$$

and (3.3.4.6) becomes

$$nE(Z) = nE(P), \qquad (3.3.4.8)$$

where the expectation is with respect to the distributions of $\mathbf{X}$ and Z; in particular, the distribution of Z is the correlated logistic. However, just as shown in the previous section

$$E(Z) = E_X(F)$$

and

$$E(\mathbf{X}Z) = E_X(\mathbf{X}F)$$

where F is the probability that Z is 1 under the correlated logistic model.

Now P is a function of Z and $\mathbf{X}$ in the sense that

$$p = \exp(\gamma_i + \boldsymbol{\beta}\mathbf{x}_{ij})/[1+\exp(\gamma_i + \boldsymbol{\beta}\mathbf{x}_{ij})]$$

where γ_i changes from cluster to cluster,

$$i=1,\ldots.$$

However, any estimate of γ_i obtained from the equations (3.3.4.4) and (3.3.4.5) is a function of $\hat{\boldsymbol{\beta}}_d$, $\mathbf{X}$ and Z (see, for example, Appendix C.3). Hence the limiting value of the estimator of γ_i is a function of $\boldsymbol{\beta}_0$ (the limiting value of $\hat{\boldsymbol{\beta}}_d$), Z and $\mathbf{X}$, in particular, the moments of Z and $\mathbf{X}$. Thus E(P) and E($\mathbf{X}$P) may be taken to denote expectation with respect to Z and $\mathbf{X}$, and the equations may be written

$$E_{Z,X}(\mathbf{X}F) = E_{Z,X}(\mathbf{X}P)$$

and

$$E_{Z,X}(F) = E_{Z,X}(P).$$

But these equations are similar to those obtained for the usual estimator $\hat{\boldsymbol{\beta}}_u$ in section (3.3.3). The arguments used in that section may also be used for the equations above. Hence the conclusions for the dummy variables estimator $\hat{\boldsymbol{\beta}}_d$ are the same as those for the usual estimator, that is,

1. when $\boldsymbol{\beta}$ is $\mathbf{0}$, $\hat{\boldsymbol{\beta}}_d$ is a consistent estimator of $\boldsymbol{\beta}$.

2. when a subset of the parameters of β is non-zero, $\hat{\beta}_d$ is a consistent estimator of the remaining parameters when the covariates corresponding to the two subsets of parameters are distributed independently of each other.

3. when the intracluster correlation is zero, $\hat{\beta}_d$ is a consistent estimator of β (although probably not as efficient as $\hat{\beta}_u$).

In order to obtain these results, we summed over one set of equations namely (3.3.4.5). The true properties of the dummy variables estimator, which satisfy those equations, namely

$$\sum_{j=1}^{n} z_{ij} = \sum_{j=1}^{n} p_{ij}$$

for each cluster are probably stronger than indicated above. In chapter 4, it is suggested that the dummy varibles estimator is a consistent estimator for β asymptotically in n, the cluster size. This implies that the dummy variables method is a reasonable approach to account for clustering when the number of elements per cluster is sufficiently large.

3.4 Asymptotic variance of estimators

3.4.1 Maximum likelihood for the correlated logistic

The log likelihood function for a single observation in a cluster of size n from the correlated logistic distribution may be written

$$L = \alpha_r + \sum_{j=1}^{n} y_j \boldsymbol{\beta}'\mathbf{x}_j - \log d,$$

where r is the number of occurrences, that is, the number of units with Y values of 1, and

$$d = \sum_{t=0}^{n} \sum_{i=1}^{{}_nC_t} \exp(\alpha_t + \sum_{j \in I_i} \boldsymbol{\beta}'\mathbf{x}_j)$$

where I_i is the i'th index set, that is, the set of the indices of the units that have Y equal to 1.

The first derivative of L with respect to α_h, h=1,...,n, may be written

$$\frac{\partial L}{\partial \alpha_h} = \delta_{hr} - \frac{1}{d}\frac{\partial d}{\partial \alpha_h},$$

where δ_{hr} is the Kronecker delta and

$$\frac{\partial d}{\partial \alpha_h} = \sum_{i=1}^{{}_nC_h} \exp(\alpha_h + \sum_{j \in I_i} \boldsymbol{\beta}'\mathbf{x}_j).$$

Hence the derivative may be written

$$\frac{\partial L}{\partial \alpha_h} = \delta_{hr} - g_h \qquad (3.4.1.1)$$

$$= {}_{n-1}C_{j-1}\, f_{12\ldots j}$$

so that the j´th element of $(\mathbf{g}_1 - f\mathbf{g})$ is

$${}_{n-1}C_{j-1}\, f_{12\ldots j}\, (j-nf)/j.$$

Now the right hand side of (3.4.1.10) may be written $h\mu_X'$, where the j´th element of h is

$${}_nC_j\, f_{12\ldots j}\, (j-nf).$$

3.4.2 Maximum likelihoood under a constrained model

Under the constraints proposed by Rosner (see Section 2.3), the n parameters, $\{\alpha_r, r=1,\ldots,n\}$, of the correlated logistic model are reduced to 2 parameters, which may be written either as α_1 and α_2, or as a and b, the parameters of the underlying beta distribution. Other parametrizations are possible, each consisting of writing the parameters, $\{\alpha_r, r=1,\ldots,n\}$, as functions of a smaller number of parameters. Let us denote the s new parameters by $\gamma_1,\ldots,\gamma_s$. We can write the original n parameters as functions of the s γ parameters, namely,

$$\alpha_j = \alpha_j(\gamma_1,\ldots,\gamma_s),\ j=1,\ldots,n;$$

for example, Rosner has proposed the functions

$$\alpha_j = \log[(\gamma_1)_j/(\gamma_2+n-j)_j],\ j=1,\ldots n,$$

where

$$(c)_j = c(c+1)\ldots(c+j-1).$$

In general, we do not constrain the β parameters.

$$X = (\mathbf{x}_1 \; \mathbf{x}_2 \; \ldots \; \mathbf{x}_n), \qquad (3.4.1.3)$$

and $\mathbf{w}_h$ as the h´th column of X´, then (3.4.1.2) may be written as

$$\frac{\partial L}{\partial \beta_h} = \mathbf{w}_h{}'(\mathbf{y}-\mathbf{f}), \; h=1,\ldots p.$$

The calculation of second derivatives proceeds in a similar manner:

$$\frac{\partial^2 L}{\partial \alpha_i \partial \alpha_h} = - \frac{\partial}{\partial \alpha_i} g_h$$

$$= - \{\frac{\partial}{\partial \alpha_i} [\sum_{j=1}^{{}_nC_h} \exp(\alpha_h + \sum_{m \in I_j} \boldsymbol{\beta}'\mathbf{x}_m)]\}/d$$

$$+ g_h \frac{\partial d}{\partial \alpha_i} /d$$

$$= - g_{ih} g_i + g_h g_i \qquad (3.4.1.4)$$

In other words we may write

$$\frac{\partial^2 L}{\partial \alpha_i \partial \alpha_h} = g_i g_h, \; i \neq h=1,\ldots,n$$

$$= - g_h (1 - g_h), \; i=h=1,\ldots,n.$$

For the derivatives with respect to β_h and α_i we have

$$\frac{\partial^2 L}{\partial \alpha_i \partial \beta_h} = - \frac{\partial}{\partial \alpha_i} \frac{\partial L}{\partial \beta_h}$$

$$= - \sum_{j=1}^{n} x_{jh} \frac{\partial}{\partial \alpha_i} f_j$$

where we write

$$\frac{\partial f_j}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} \sum_{t=1}^{n} \sum_{m=1}^{{}_{n-1}C_{t-1}} \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}'\mathbf{x}_p) /d$$

$$= \sum_{t=1}^{n} \sum_{m=1}^{{}_{n-1}C_{t-1}} \frac{\partial}{\partial \alpha_i} \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}'\mathbf{x}_p) /d$$

$$- f_j \frac{\partial d}{\partial \alpha_i} / d$$

$$= \sum_{m=1}^{{}_{n-1}C_{i-1}} \exp(\alpha_i + \sum_{p \in I_{mj}} \boldsymbol{\beta}' \mathbf{x}_p) - f_j g_i$$

$$= g_{ij} - g_i f_j, \qquad (3.4.1.5)$$

where g_{ij} is the joint probabinildty that i of the units have a Y value of 1 and that one of these is unit j. Hence the second derivative may be written

$$\frac{\partial^2 L}{\partial \alpha_i \partial \beta_h} = - \sum_{j=1}^{n} x_{ih} (g_{ij} - g_i f_j), \quad i=1,\ldots,n, \quad h=1,\ldots p.$$

Note that g_i and f_j are the marginal probabilities whereas g_{ij} is the joint probability.

The second derivative with respect to β_i and β_h may be written

$$\frac{\partial^2 L}{\partial \beta_i \partial \beta_h} = - \sum_{j=1}^{n} x_{jh} \frac{\partial}{\partial \beta_i} f_j$$

where

$$\frac{\partial f_j}{\partial \beta_i} = \sum_{t=1}^{n} \sum_{m=1}^{{}_{n-1}C_{t-1}} [\frac{\partial}{\partial \beta_i} \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}' \mathbf{x}_p)]/d$$

$$- f_j \frac{\partial d}{\partial \beta_i} / d$$

$$= \sum_{t=1}^{n} \sum_{m=1}^{{}_{n-1}C_{t-1}} [(\sum_{p \in I_{mj}} x_{pi}) \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}' \mathbf{x}_p)] / d$$

$$- f_j \sum_{t=1}^{n} x_{ti} f_t \qquad (3.4.1.6)$$

For the first term in this expression consider the coefficient of a typical x factor, say x_{ti}. Now x_{ti} appears only when t is in the index set I_{mj}, that is, when x_t appears in the term

$$\exp(\alpha_h + \Sigma\, \boldsymbol{\beta}'\mathbf{x}_p).$$

When this occurs, we are calculating a probability that contains the event

$$Y_t = 1.$$

Thus the coefficient of x_{ti} in the first term of expression (3.4.1.6) is the probability that Y_t is 1 and Y_j is 1; that is, (3.4.1.6) may be written

$$= \sum_{t=1}^{n} x_{ti}\, f_{tj} - f_j \sum_{t=1}^{n} x_{ti}\, f_t$$

where

$$f_{tj} = \Pr(Y_t = 1, Y_j = 1).$$

and, when

$$t = j$$

then

$$f_{tj} = f_t = f_j = \Pr(Y_j = 1) = \Pr(Y_t = 1).$$

Hence $\partial^2 L/\partial\beta_i \partial\beta_h$ may be written

$$- \sum_{j=1}^{n} (x_{jh} \sum_{t=1}^{n} x_{ti}\ f_{jt}) + (\sum_{j=1}^{n} x_{jh}\ f_j)(\sum_{t=1}^{n} x_{ti}\ f_t),$$

which may be rewritten as

$$- \sum_{j=1}^{n} [x_{jh}\ x_{ji}\ f_j(1-f_j)] - \sum_{j\neq t}^{n} [x_{jh}\ x_{ti}\ (f_{jt} - f_j f_t)].$$

In summary the matrix of minus the second derivatives may be written as

$$\begin{bmatrix} W - \mathbf{gg}' & (G-\mathbf{gf}')X' \\ X(\mathbf{fg}'-G') & X(F-\mathbf{ff}')X' \end{bmatrix} \tag{3.4.1.7}$$

where W is the diagonal matrix of the g_i's, $i=1,\ldots,n$, **g** is the vector of the g_i's, and

$$g_i = \Pr(i \text{ of the } Y\text{'s are } 1, \text{ the remainder are } 0),$$

f is the vector of the f_i's, $i=1,\ldots n$, f_i is the marginal probability that Y_i is 1, that is,

$$f_i = \Pr(Y_i = 1),$$

G is the matrix of g_{ij}'s, $i,j=1,\ldots n$, where

$$g_{ij} = \Pr(Y_j=1, \text{ as are } i-1 \text{ of the other } Y\text{'s}),$$

F is the matrix of the f_{ij}'s, $i,j=1, . n$, where

$$f_{ij} = \Pr(Y_i = Y_j = 1),$$

f_{ii} also being known as f_i, and X is the matrix of covariates as defined in (3.4.1.4).

The inverse of the expectation of this matrix yields the asymptotic dispersion matrix of the maximum likelihood estimator of the correlated logistic model. This matrix is studied further in the next section in which constraints are imposed on the correlated model, yielding Rosner's correlated logistic model.

The expectation of the matrix (3.4.1.7) simplifies somewhat if we make the assumption that the marginal distribution of **X** is the same in all the units and that the joint distribution of any pair of X's is the same for any pair of units. In particular,

$$E_X[X(F-\mathbf{ff}')X']$$

$$= n\, E_X[\mathbf{XX}' f(1-f) + (n-1)\, \mathbf{X}_1\mathbf{X}_2'(f_{12}-f_1 f_2)], \quad (3.4.1.8)$$

where $\mathbf{XX}'$ indicates that observations are from the same unit, and $\mathbf{X}_1\mathbf{X}_2'$ indicates that observations are from two different units. Similarly,

$$f = Pr(Y=1),$$

$$f_i = Pr(Y_i = 1), \; i=1,2,$$

and

$$f_{12} = Pr(Y_1 = 1, Y_2 = 1).$$

A further simplification occurs when the regression parameter is zero, that is, when the covariates have no effect. In the first place, f, f_1, f_2 and f_{12} are no longer functions of X and may be removed from within the expectation brackets. If we denote f(1-f) as σ_y^2, then

$$f_{12}-f_1 f_2 = f_{12} - f^2$$
$$= \rho_y \, \sigma_y^2,$$

so that (3.4.1.8) becomes

$$n\, \sigma_y^2\, [E(\mathbf{XX}') + (n-1)\rho_y\, E(\mathbf{X}_1\mathbf{X}_2')]. \quad (3.4.1.9)$$

Also

$$E[(G-\mathbf{gf}')X'] = n\, (\mathbf{g}_1 - f\mathbf{g})\boldsymbol{\mu}_X', \quad (3.4.1.10)$$

where f and $\mathbf{g}$ have been defined above, but now g_j, the j'th element of $\mathbf{g}$, is equal to

$${}_nC_j \, f_{12\ldots j};$$

where

$$f_{12\ldots j} = Pr(Y_1=1, \ldots Y_j=1, Y_{j+1}=0, \ldots Y_n=0),$$

and $\mathbf{g}_1$ is the vector whose j'th element, g_{1j}, is defined as

$$g_{1j} = Pr(Y_1=1, \text{ as do } (j-1) \text{ of the other } Y\text{'s})$$

$$= {}_{n-1}C_{j-1}\ f_{12\ldots j}$$

so that the j'th element of $(\mathbf{g}_1 - f\mathbf{g})$ is

$$ {}_{n-1}C_{j-1}\ f_{12\ldots j}\ (j-nf)/j.$$

Now the right hand side of (3.4.1.10) may be written $\mathbf{h}\boldsymbol{\mu}_X'$, where the j'th element of h is

$$ {}_{n}C_{j}\ f_{12\ldots j}\ (j-nf).$$

3.4.2 Maximum likelihoood under a constrained model

Under the constraints proposed by Rosner (see Section 2.3), the n parameters, $\{\alpha_r, r=1,\ldots,n\}$, of the correlated logistic model are reduced to 2 parameters, which may be written either as α_1 and α_2, or as a and b, the parameters of the underlying beta distribution. Other parametrizations are possible, each consisting of writing the parameters, $\{\alpha_r, r=1,\ldots,n\}$, as functions of a smaller number of parameters. Let us denote the s new parameters by $\gamma_1,\ldots,\gamma_s$. We can write the original n parameters as functions of the s γ parameters, namely,

$$\alpha_j = \alpha_j(\gamma_1,\ldots,\gamma_s),\ j=1,\ldots,n;$$

for example, Rosner has proposed the functions

$$\alpha_j = \log[(\gamma_1)_j/(\gamma_2+n-j)_j],\ j=1,\ldots n,$$

where

$$(c)_j = c(c+1)\ldots(c+j-1).$$

In general, we do not constrain the β parameters.

Since the β parameters are not changed, we need only investigate the effects of constraints on the derivatives with respect to the α parameters. For the first derivative with respect to γ_i, we have

$$\frac{\partial L}{\partial \gamma_i} = \frac{\partial A_r}{\partial \gamma_i} - \frac{1}{d}\,\frac{\partial d}{\partial \gamma_i}$$

where A_r is a random variable in the sense that the actual function of γ_i that appears depends on the value of **Y**, the outcome vector. This point will be discussed later in this section. The right hand side of the above equation becomes

$$\frac{\partial A_r}{\partial \gamma_i} - \sum_{j=1}^{n} g_j \frac{\partial \alpha_j}{\partial \gamma_i}, \quad i=1,\ldots s,$$

where g_j was defined in the previous section as

g_j = Pr(j of the Y's have value 1).

The second derivatives also change from those given in the previous section, so that

$$\frac{\partial^2 L}{\partial \gamma_h \partial \gamma_i} = \frac{\partial^2 A_r}{\partial \gamma_h \partial \gamma_i} - \sum_{j=1}^{n} \left[\frac{\partial}{\partial \gamma_h} \left(g_j \frac{\partial \alpha_j}{\partial \gamma_i} \right) \right]$$

$$= \frac{\partial^2 A_r}{\partial \gamma_h \partial \gamma_j} - \sum_{j=1}^{n} \left(\frac{\partial \alpha_j}{\partial \gamma_i} \frac{\partial g_j}{\partial \gamma_j} + g_j \frac{\partial^2 \alpha_j}{\partial \gamma_h \partial \gamma_i} \right)$$

$$= \frac{\partial^2 A_r}{\partial \gamma_h \partial \gamma_i} - \sum_{j=1}^{n} \left(g_j \frac{\partial^2 \alpha_j}{\partial \gamma_h \partial \gamma_i} \right) - \sum_{j=1}^{n} \left(\frac{\partial \alpha_j}{\partial \gamma_i} \frac{\partial \gamma_j}{\partial \gamma_h} \right)$$

Now

$$\frac{\partial \gamma_j}{\partial \gamma_h} = \left\{ \sum_{t=1}^{{}_nC_t} \left[\frac{\partial}{\partial \gamma_h} \exp(\alpha_j + \sum_{m \in I_t} \boldsymbol{\beta}' \mathbf{x}_m) \right] - g_j \frac{\partial d}{\partial \gamma_h} \right\} / d$$

$$= g_j \frac{\partial \alpha_j}{\partial \gamma_h} - \sum_{t=1}^{n} g_t \frac{\partial \alpha_t}{\partial \gamma_h} .$$

Hence $\partial^2 L/\partial\gamma_h\partial\gamma_i$ may be written

$$\frac{\partial^2 A_r}{\partial\gamma_h\partial\gamma_i} - \sum_{j=1}^{n} g_j \frac{\partial^2\alpha_j}{\partial\gamma_h\partial\gamma_i} - \sum_{j=1}^{n} g_j \frac{\partial\alpha_j}{\partial\gamma_h}\frac{\partial\alpha_j}{\partial\gamma_i} + \sum_{j=1}^{n} g_j \frac{\partial\alpha_j}{\partial\gamma_h} \sum_{j=1}^{n} g_j \frac{\partial\alpha_j}{\partial\gamma_i} .$$

Consider the first two terms in this expression, namely,

$$\frac{\partial^2 A_r}{\partial\gamma_h\partial\gamma_i} - \sum_{j=1}^{n} g_j \frac{\partial^2\alpha_j}{\partial\gamma_h\partial\gamma_i} .$$

It is obvious that, for a full set of n α parameters, this term is zero, but of interest in this section is the value of this expression when the number of parameters is less than n. In particular, this expression will be used in obtaining the asymptotic variance of the maximum likelihood estimators. The form of the asymptotic dispersion matrix may be simplified by the following lemma.

Lemma 3.4.2.

$$E_{Y|X}\left(\frac{\partial^2 A_r}{\partial\gamma_h\partial\gamma_i} \right) = \sum_{j=1}^{n} g_j \frac{\partial^2\alpha_j}{\partial\gamma_h\partial\gamma_i} .$$

Proof: A_r is a random variable that takes on values α_j with probability g_j, where g_j was defined above. Hence

$$E(A_r) = \sum_{j=1}^{n} \alpha_j g_j .$$

Moreover, $\partial^2 A_r/d\gamma_h d\gamma_i$ is a random variable that takes on values $\partial^2\alpha_j/d\gamma_h d\gamma_i$ with probability g_j, $j=1,\ldots,n$, so that

$$E\left(\frac{\partial^2 A}{\partial\gamma_h \partial\gamma_i}\right) = \sum_{j=1}^{n} \left(\frac{\partial^2 \alpha_j}{\partial\gamma_h \partial\gamma_i} g_j \right) .$$

That A_r is a random variable can be seen from the fact that the use of α_r in the distribution function of the correlated logistic is merely a short-hand notation for the expression

$$\delta_1 \sum_{j=1}^{n} y_j + \delta_2 \sum_{j \neq m}^{n} y_j y_m + \ldots + \delta_n \prod_{j=1}^{n} y_j ,$$

in the Dempster generalised exponential family (see Section 2.2), and the particular function of Y used with α_r is that function which takes the value 1 when r of the Y's are 1, and otherwise is zero.

The second derivative $\partial^2 L/\partial\alpha_h \partial\beta_i$ may be evaluated as

$$-\frac{\partial}{\partial\gamma_h}\frac{\partial L}{\partial\beta_i} = -\sum_{j=1}^{n} x_{ji} \frac{\partial f_j}{\partial\gamma_h}$$

Now $\partial f_j/\partial\gamma_h$ is different from the expression for $\partial f_j/\partial\alpha_h$ given in (3.4.1.5), in particular,

$$\frac{\partial f_j}{\partial\gamma_h} = \frac{\partial}{\partial\gamma_h} \left[\sum_{t=1}^{n} \sum_{m=1}^{{}_{n-1}C_{t-1}} \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}' \mathbf{x}_p)/d \right]$$

$$= \sum_{t=1}^{n} \left[\sum_{m=1}^{{}_{n-1}C_{t-1}} \exp(\alpha_t + \sum_{p \in I_{mj}} \boldsymbol{\beta}' \mathbf{x}_p) \frac{\partial\alpha_t}{\partial\gamma_h} \right]/d$$

$$- f_j \frac{\partial d}{\partial\gamma_h} /d$$

$$= \sum_{t=1}^{n} g_{tj} \frac{\partial\alpha_t}{\partial\gamma_h} - f_j \sum_{t=1}^{n} g_{t\cdot} \frac{\partial\alpha_t}{\partial\gamma_h} .$$

Thus we have

$$\frac{\partial^2 L}{\partial\gamma_h \partial\beta_i} = - \sum_{j=1}^{n} \{x_{ji} [\sum_{t=1}^{n} (g_{tj} - f_j g_t) \frac{\partial\alpha_t}{\partial\gamma_h}]\},$$

$$h=1,\ldots,s,\ i=1,\ldots p.$$

From the results in this section, it may be seen that, for any reduced form of the correlated logistic distribution, and in particular for Rosner's model, the asymptotic variance-covariance matrix is calculated from the inverse of the expectation of a matrix similar to that given in (3.4.1.7), but with

$$W - \mathbf{gg}'$$

replaced by the matrix

$$HWH' - H\mathbf{gg}'H'$$

and

$$(G - \mathbf{gf}')X'$$

replaced by the matrix

$$H(G - \mathbf{gf}')X',$$

where H is the (s×n) matrix $\partial\boldsymbol{\alpha}/\partial\boldsymbol{\gamma}$, that is, the matrix whose (i,j) element is $\partial\alpha_j/\partial\gamma_i$.

3.4.3 Conditional estimator

As shown in section 2.4, the conditional probability distribution function for a single cluster can be written

$$\frac{\exp(\sum_{j\epsilon I_1} \boldsymbol{\beta}'\mathbf{x}_j)}{\sum_{i=1}^{{}_nC_r} \exp(\sum_{j\epsilon I_i} \boldsymbol{\beta}'\mathbf{x}_j)}$$

where I_1 is the index set for the particular sample observed, that is, the set of indices $\{j_1,\ldots,j_r\}$ of the Y's with value 1, and $\{I_i, i=1,\ldots {}_nC_r\}$ is the set of all possible index sets having exactly r Y values equal to 1. Without loss of generality, we may assume that the first r units have Y equal to 1 (or that the units have been relabelled to produce such a result) so that the distribution function for a single cluster may be written

$$p(\sum_{j=1}^{r} \mathbf{x}_j) = \exp(\sum_{j=1}^{r} \boldsymbol{\beta}'\mathbf{x}_j) / d \qquad (3.4.3.1)$$

where

$$d = \sum_{i=1}^{{}_nC_r} \exp(\sum_{j\epsilon I_i} \boldsymbol{\beta}'\mathbf{x}_j).$$

Expression (3.4.3.1) is also the likelihood function for β, so that the conditional log likelihood function is

$$L(\boldsymbol{\beta}) = \sum_{j=1}^{r} \boldsymbol{\beta}'\mathbf{x}_j - \log d. \qquad (3.4.3.2)$$

The first derivative of $L(\boldsymbol{\beta})$ with respect to β_i, an element of β, is

$$\sum_{j=1}^{n} x_{ji} - \frac{1}{d}\frac{\partial d}{\partial \beta_i},$$

where $\partial d/\partial\beta_i$ is

$$\sum_{j=1}^{{}_nC_r} \sum_{m\epsilon I_j} x_{mi} \exp(\sum_{m\epsilon I_j} \boldsymbol{\beta}'\mathbf{x}_m).$$

so that

$$\frac{1}{d}\frac{\partial d}{\partial \beta_i} = \sum_{j=1}^{{}_nC_r} \sum_{m\in I_j} x_{mi}\; p(\sum_{m\in I_j} \mathbf{x}_m)$$

where $p(\sum_{i\in I_j} \mathbf{x}_m)$ was defined as the conditional probability distribution function in (3.4.3.1). Hence the first derivative with respect to β may be written

$$\frac{\partial L}{\partial \beta_i} = \sum_{j=1} x_{ji} - \sum_{j=1}^{{}_nC_r} \sum_{m\in I_j} x_{mi}\; p(\sum_{m\in I_j} \mathbf{x}_m),$$

$$i=1,\ldots,p.$$

The second derivatives may be determined in a similar way, that is,

$$\frac{\partial^2 L}{\partial \beta_h \partial \beta_i} = -\sum_{j=1}^{{}_nC_r} \sum_{m\in I_j} x_{mi} \frac{dp(\Sigma \mathbf{x}_m)}{\partial \beta_h},$$

where the summation counter has been omitted from $p(\Sigma \mathbf{x}_m)$. Now $\partial p(\Sigma \mathbf{x}_m)/\partial \beta_h$ may be written as

$$\frac{\partial}{\partial \beta_h} \frac{\exp(\Sigma\, \boldsymbol{\beta}'\mathbf{x}_m)}{d} = \frac{\partial}{\partial \beta_h} [\exp(\sum_{m\in I_j} \boldsymbol{\beta}'\mathbf{x}_m)] /d$$

$$- \exp(\sum_{m\in I_j} \boldsymbol{\beta}'\mathbf{x}_m) \frac{\partial d}{\partial \beta_h} /d$$

$$= \sum_{m\in I_j} x_{mh}\; p(\sum_{m\in I_j} \mathbf{x}_m) - p(\sum_{m\in I_j} \mathbf{x}_m) \sum_{m=1}^{{}_nC_r} \sum_{p\in I_m} x_{ph}\; p(\sum_{p\in I_m} \mathbf{x}_p)$$

so that

$$\frac{\partial^2 L}{\partial \beta_h \partial \beta_i} = -\sum_{j=1}^{{}_nC_r} [(\sum_{m\in I_j} x_{mh})(\sum_{m\in I_j} x_{mi}) p(\sum_{m\in I_j} \mathbf{x}_m)]$$

$$+ \left[\sum_{j=1}^{{}_nC_r} \left(\sum_{m \in I_j} x_{mh} \right) p\left(\sum_{m \in I_j} \mathbf{x}_m \right) \right] \left[\sum_{j=1}^{{}_nC_r} \left(\sum_{m \in I_j} x_{mi} \right) p\left(\sum_{m \in I_j} \mathbf{x}_m \right) \right] \qquad (3.4.3.2)$$

$$h, i = 1, \ldots, p.$$

The information matrix for the conditional correlated logistic distribution is given by

$$E\left(- \frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right)$$

where the (h,i)'th element of $\partial^2 L / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'$ is given by (3.4.3.2).

Since the distribution of $\mathbf{X}$ is assumed to be symmetric with respect to the units, the expectation of the (h,i)'th element of the information matrix simplifies to

$${}_nC_r \left\{ E\left[\left(\sum_{m=1}^{r} x_{mh} \right) \left(\sum_{m=1}^{r} x_{mi} \right) p\left(\sum_{m=1}^{r} \mathbf{x}_m \right) \right] \right.$$

$$\left. - \sum_{j=1}^{{}_nC_r} E\left[\left(\sum_{m=1}^{r} x_{mh} \right) p\left(\sum_{m=1}^{r} \mathbf{x}_m \right) \left(\sum_{m \in I_j} x_{mh} \right) p\left(\sum_{m \in I_j} \mathbf{x}_m \right) \right] \right\} \qquad (3.4.3.3)$$

This expression will be examined in some special cases in Chapter 4.

In the null case, that is, when

$$\boldsymbol{\beta} = \mathbf{0},$$

the probability distribution function (3.4.3.1) simplifies to

$$p(\cdot) = 1/{}_nC_r$$

so that (3.4.3.3) becomes

$$E\left(\sum_{m=1}^{r} x_{mh} \sum_{m=1}^{r} x_{mi} \right) - \left[\sum_{j=1}^{{}_nC_r} E\left(\sum_{m=1}^{r} x_{mh} \sum_{m \in I_j} x_{mi} \right) \right] / {}_nC_r \qquad (3.4.3.4)$$

The first term of this expression may be written

$$r\,E(X_h X_i) + r(r-1)\,E(X_{h1} X_{i2}) \qquad (3.4.3.5)$$

where $X_h X_i$ indicates that the variables are measures on the same unit (and hence have a within-unit but not a between-unit component). Similarly $X_{h1} X_{i2}$ indicates that the variables are measured on different units within the same cluster (and hence have a between-unit component).

The evaluation of the second term in expression (3.4.3.5) is more complicated. Let us consider the expression

$$\left(\sum_{m=1}^{r} x_{mh}\right)\left(\sum_{m \varepsilon I_j} x_{mi}\right)$$

Assume that t of the indices in I_j are also in the set $\{1,2,\ldots,r\}$ so that

$$E\left[\left(\sum_{m=1}^{r} x_{mh}\right)\left(\sum_{m \varepsilon I_j} x_{mi}\right)\right] = tE(X_h X_i) + (r^2 - t)E(X_{h1} X_{i2})$$

and

$$\sum_{j=1}^{{}_nC_r} \left\{E\left[\left(\sum_{m=1}^{r} x_{mh}\right)\left(\sum_{m \varepsilon I_j} x_{mi}\right)\right]\right\} / {}_nC_r$$

$$= \sum_{t=0}^{r} {}_rC_t \; {}_{n-r}C_{r-t} \; [tE(X_h X_i)$$

$$+ (r^2 - t)E(X_{h1} X_{i2})] / {}_nC_r \qquad (3.4.3.6)$$

but

$$\frac{{}_rC_t \; {}_{n-r}C_{r-t}}{{}_nC_r}$$

is the probability distribution function of the hypergeometric distribution with population of size n, of which r have the required characteristic and a sample of size r is picked. Thus T may be considered as a random variable with a hypergeometric distribution and (3.4.3.6) may be written

$$E(T)\, EX(X_h X_i) + [r^2 - E(T)] EX(X_{h1} X_{i2})$$
$$= [r^2 E(X_h X_i) + (nr^2 - r^2) E(X_{h1} X_{i2})]/n$$
$$= r^2 [E(X_h X_i) + (n-1) E(X_{h1} X_{i2})]/n$$

Thus (3.4.3.4) becomes

$$r(n-r)\, [E(X_h X_i) - E(X_{h1} X_{i2})]/n,$$

In general, the information matrix may be written

$$E(-\frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}) = r(n-r)\, [E(\mathbf{XX}') - E(\mathbf{X}_1 \mathbf{X}_2')]/n, \quad (3.4.3.7)$$

where the diagonal elements are of the form

$$r(n-r)\, [E(X_i^2) - E(X_{i1} X_{i2})]/n$$
$$= r(n-r)\, \sigma^2_{X_i}\, (1 - \rho_{X_i}),$$

where ρ_{X_i} is the intraclass (or intracluster) correlation for variable X_i.

The expressions above give the conditional information, that is, the information in a single cluster conditional on there being exactly r units with Y equal to 1 and the remaining (n-r) Y's equal to 0. In order to make comparisons with other estimators, we require an unconditional matrix corresponding to information over all sample sizes. We shall obtain this information matrix by "unconditioning" the conditional matrix, that is, by taking the expectation with respect to all possible values of Y. This is given by

$$E_{Y,X}\left(-\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right)$$

$$= \sum_{r=0}^{n} E_X\left[\left(-\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right) \Pr(r\ Y's\ \text{are}\ 1)\right] \quad (3.4.3.8)$$

where the typical element in $\partial^2 L/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'$ is given by (3.4.3.2). Pr(r Y´s are 1) is the conditional probability according to the correlated logistic distribution, and is generaly dependent on the values of X. Expression(3.4.3.8) is evaluated for some simple cases in Chapter 4.

However, when

$$\boldsymbol{\beta} = \mathbf{0},$$

the probability, Pr(r Y´s are 1), does not depend on X and the expression for the information simplifies, in particular, from expression (3.4.3.7) we get

$$E\left(-\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right)$$

$$= \sum_{r=0}^{n} r(n-r)[E(\mathbf{XX}')-E(\mathbf{X}_1\mathbf{X}_2')]\ g_r/n, \qquad (3.4.3.9)$$

This expression uses the term g_r, which was defined in Section 3.4.1 as

$$g_r = \Pr(r \text{ Y's are 1 and the remaining } (n-r) \text{ Y's are 0})$$

$$= {}_nC_r \, h_r$$

where

$$h_r = \Pr(\text{a particular set of r Y's are 1, the rest being 0}).$$

For example, under the correlated logistic model,

$$h_r = \exp(\alpha_r) \Big/ \sum_{j=0}^{n} \exp(\alpha_j).$$

Note that the summation in the above expression ranges from 0 to n. Although a cluster with 0 or n Y's equal to 1 contains no information about β, that is, the likelihood is constant, this result is contained in expression (3.4.3.7) which has value 0 for

$$r=0$$

and

$$r=n.$$

Hence the expression for the information matrix becomes

$$[E(\mathbf{XX}') - E(\mathbf{X}_1\mathbf{X}_2')] \sum_{r=0}^{n} \frac{r(n-r)}{n} g_r \qquad (3.4.3.10)$$

As indicated in Chapter 2, when

$$\boldsymbol{\beta} = \mathbf{0},$$

$\{g_r, r=0,\ldots,n\}$ is the probability distribution function of the beta-binomial distribution. Let R denote the beta-binomial random variable. The summation expression in (3.4.3.10) may be written

$$\sum_{r=0}^{n} \frac{r(n-r)}{n} g_r$$

$$= E(R) - E(R^2)/n \qquad (3.4.3.11)$$

The beta-binomial distribution of R may be written in terms of parameters n, a and b, or, as indicated by Crowder(1979) and Williams(1982), in terms of parameters n, θ and ρ, where

$$\theta = a/(a+b)$$

and

$$\rho = 1/(a+b+1)$$

where ρ has been denoted earlier as ρ_y, the intraclass (intracluster) coefficient of Rosner's correlated logistic model, that is,

$$\rho_y = \text{corr}(Y_i, Y_j), \quad i,j=1,\ldots n.$$

Under this parametrization,

$$E(Y) = \theta$$

and

$$\text{Var}(Y) = \theta(1-\theta) = \sigma_y^2.$$

Also

$$E(R) = n\theta$$

and

$$\text{Var}(R) = n\theta(1-\theta)[1+\rho_y(n-1)].$$

We may write the expression (3.4.3.11) as

$$n\theta - \{n\theta(1-\theta)[1+\rho_y(n-1)]+(n\theta)^2\}/n$$

which, after some manipulation, becomes

$$(n-1)\ \sigma_y^2(1-\rho_y).$$

Thus, under Rosner's correlated logistic model, with

$$\boldsymbol{\beta} = \mathbf{0},$$

the expected information from the conditional model can be written

$$I(\beta) = (n-1)\, \sigma_y^2 (1-\rho_y)\, [E(\mathbf{XX}') - E(\mathbf{X}_1\mathbf{X}_2')]. \quad (3.4.3.12)$$

3.4.4 Usual estimator

The term 'usual estimator' denotes the estimator derived by solving the maximum likelihood equations for the usual (or uncorrelated) logistic regression model. We wish to study the distribution of this estimator when the assumption of no correlation is wrong, that is, when the correlated logistic distribution is correct. A general method for examining the distribution of an estimator, $\hat{\boldsymbol{\theta}}$, that has been obtained by solving the estimating equation(s)

$$S_k(\hat{\boldsymbol{\theta}}) = \mathbf{0},$$

where k denotes the sample size, is given is Appendix B.4. We have investigated the expected value of S_k in Section 3.3.3. In this section, we are interested in the asymptotic variance of $\hat{\boldsymbol{\theta}}$, which is shown in the appendix to be $U^{-1}VU^{-1}$, where V is the asymptotic variance of S_k and U^{-1} is the limiting value of $[-\mathbf{S}_k'(\hat{\boldsymbol{\theta}})/k]$ where $\mathbf{S}_k'$ is the derivative of $\mathbf{S}_k$ with respect to the parameter, $\boldsymbol{\theta}$.

We are also interested in the misspecifiation effect, that is, the error in estimating the variance of $\hat{\boldsymbol{\theta}}$ due to the misspecification of the probability distribution. The usual estimator of the variance of $\hat{\boldsymbol{\theta}}$ is $(-\mathbf{S}_k'/k)$, which, as indicated above, approaches U^{-1}. Thus the misspecification factor is VU^{-1}. This misspecifiaction effect is analagous to that defined by Scott and Holt(1982) for the case of continuous dependent variables.

For the usual estimator of the regression coefficient, denoted here by $\hat{\boldsymbol{\beta}}_u$, it has been shown in section 3.3.1 that $\hat{\boldsymbol{\beta}}_u$ estimates $\boldsymbol{\beta}$ consistently when $\boldsymbol{\beta}$ is zero. Hence we are only investigating the asymptotic variance of $\hat{\boldsymbol{\beta}}_u$ when $\boldsymbol{\beta}$ is zero.

If one were interested in measuring behaviour of an estimator in terms of mean square error, then one might want to evaluate the asymptotic variance of $\hat{\boldsymbol{\beta}}_u$ when $\boldsymbol{\beta}$ was non-zero. However, that is not done in this thesis.

The estimating equations, $\mathbf{S}_k$, used for obtaining the usual estimator are those of maximum likelihood estimation under the usual (or uncorrelated) logistic model, that is, for a log likelihood

$$\alpha \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij} + \sum_{i=1}^{k} \sum_{j=1}^{n} \boldsymbol{\beta}'\mathbf{x}_{ij}\, y_{ij} - \sum_{i=1}^{k} \sum_{j=1}^{n} \log d_{ij}$$

where

$$d_{ij} = 1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x}_{ij}).$$

One takes derivatives with respect to α and $\boldsymbol{\beta}$, and sets them equal to zero, giving the equations

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - p_{ij}) = 0$$
$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - p_{ij})\mathbf{x}_{ij} = 0 \qquad (3.4.4.1)$$

where

$$p_{ij} = \Pr(Y_{ij}=1|\mathbf{x}_{ij})$$
$$= \exp(\alpha + \boldsymbol{\beta}'\mathbf{x}_{ij}) \;/\; [1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x}_{ij})].$$

It was shown in section (3.3.1) that

$$\boldsymbol{\beta} = \mathbf{0}$$

implies that

$$\hat{\boldsymbol{\beta}}_u \xrightarrow{p} 0$$

and

$$\hat{\alpha}_u \xrightarrow{p} \text{logit}(F),$$

where

$$F = \Pr(Y=1|\boldsymbol{\beta}=\mathbf{0}) \qquad (3.4.4.2)$$

where the probability is calculated under the correlated logistic model. Since Y is a binary variable, (3.4.4.2) implies that

$$E(Y) = F.$$

If we write α_0 and $\boldsymbol{\beta}_0$ as the limiting values af $\hat{\alpha}_u$ and $\hat{\boldsymbol{\beta}}_u$ then P_0, the limiting value of p_{ij} is

$$\exp(\alpha_0) \;/\; [1 + \exp(\alpha_0)] = F$$

so that

$$E(Y) = F = P_0. \qquad (3.4.4.3)$$

Equations (3.4.4.1) yield the estimating equations, that is, $\mathbf{S}_k(\alpha,\boldsymbol{\beta})$, namely,

$$\begin{aligned} & k\, E[\sum_{j=1}^{n} (Y_j - P_j)] \\ & k\, E[\sum_{j=1}^{n} \mathbf{X}_j (Y_j - P_j)] \end{aligned} \qquad (3.4.4.4)$$

Examining these estimating equations, we see that $E[\mathbf{S}_k(\alpha_0,\boldsymbol{\beta}_0)]$ may be written

$$\begin{aligned} & kn\, E(Y-P_0) \\ & kn\, E[\mathbf{X}(Y-P_0)] \end{aligned}$$

that is,

$$\begin{aligned} & kn\, E(Y-F) \\ & kn\, E[\mathbf{X}(Y-F)] \end{aligned} \qquad (3.4.4.5)$$

By equation (3.4.4.3), the first line of (3.4.4.5) is 0. Also, we may write $E[\mathbf{X}(Y-F)]$ as

$$\begin{aligned} E_X[\mathbf{X}\, E_{Y|X}(Y-F)] &= E_X(\mathbf{X} \times 0) \\ &= \mathbf{0}, \end{aligned}$$

that is,

$$E[\mathbf{S}_k(\alpha_0,\boldsymbol{\beta}_0)] = 0.$$

Moreover, $\mathbf{S}_k(\alpha_0,\boldsymbol{\beta}_0)$ is the sum of k independent, identically distributed terms, each of which has mean $\mathbf{0}$ and variance V (= $\mathrm{Var}[\mathbf{S}_k(\alpha_0,\boldsymbol{\beta}_0)]$). By the multivariate central limit theorem, it may be shown that

$$\frac{1}{\sqrt{k}}\, \mathbf{S}_k(\alpha_0,\boldsymbol{\beta}_0) \xrightarrow{D} MVN(\mathbf{0},V).$$

The matrix V may be partitioned into four submatrices corresponding to the partition of $\mathbf{S}_k(\cdot)$ into two expressions (as in (3.4.4.4)). Let V be written as

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

where

$$V_{11} = \mathrm{Var}[\sum_{j=1}^{n} (Y_j - F)]$$

$$= \mathrm{Var}(\sum_{j=1}^{n} Y_j), \text{ (since } E(Y_j) = F)$$

$$= n\,\mathrm{Var}(Y) + n(n-1)\,\mathrm{cov}(Y_1, Y_2)$$

$$= n\,\sigma_y^2\,[1 + (n-1)\rho_y].$$

To evaluate

$$V_{22} = \mathrm{Var}[\sum_{j=1}^{n} \mathbf{X}_j (Y_j - F)]$$

we require the rule

$$\mathrm{Var}_{X,Y} = E_X\,\mathrm{Var}_{Y|X} + \mathrm{Var}_X\,E_{Y|X} \quad (3.4.4.6)$$

Since

$$E(Y_j) = F, \; j=1,\ldots,n,$$

The second term of (3.4.4.6) disappears so that

$$V_{22} = E_X\,V_{Y|X}\,[\sum_{j=1}^{n} \mathbf{X}_j (Y_j - F)],$$

where $V_{Y|X}$ denotes $\mathrm{Var}_{Y|X}$. The i'th diagonal element of the matrix V_{22} is

$$E_X V_{Y|X} \sum_{j=1}^{n} X_{ij} (Y_j - F)$$

$$= n \sigma_y^2 E(X_i^2) + n(n-1) \operatorname{cov}(Y_1, Y_2) E(X_{i1}, X_{i2})$$

where X_{i1} and X_{i2} describe the correlated observations on a covariate for two units in the same cluster. The (h,i)'th element $(h \neq i)$ of matrix V_{22} may be written as

$$E_X E_{Y|X} \{ \sum_{j=1}^{n} [X_{hj} (Y_j - F)] \sum_{j=1}^{n} [X_{ij} (Y_j - F)] \}$$

$$= n \sigma_y^2 E(X_h, X_i) + n(n-1) \operatorname{cov}(Y_1, Y_2) E(X_{h1}, X_{i2}),$$

where X_h and X_i are the random variables representing two covariates and X_{h1} and X_{i2} are the random variables for the correlated observations on two units in the same cluster. Hence we may write V_{22} as

$$n \sigma_y^2 [E(\mathbf{X}\mathbf{X}') + (n-1) \rho_y E(\mathbf{X}_1 \mathbf{X}_2')],$$

where the notation was introduced in the previous section, that is, $\mathbf{X}$ is the random vector for the p covariates in a single unit in a cluster, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are the random vectors for the p covariates on two units in the same cluster.

Finally we may write V_{21} as

$$\operatorname{cov}(\sum_{j=1}^{n} Y_j, \sum_{j=1}^{n} X_j' Y_j)$$

$$= E_X \{ \sum_{j=1}^{n} \mathbf{X}_j [\sigma_y^2 + (n-1) \operatorname{cov}(Y_1, Y_2)]\}$$

$$= n \sigma_y^2 E\{X [1 + (n-1)\rho_y]\}$$

$$= n \sigma_y^2 \mu_{\mathbf{X}} [1 + (n-1)\rho_y].$$

Thus V may be written

$$n\sigma_y^2 \left[\begin{pmatrix} 1 & \mu_{\mathbf{X}}' \\ \mu_{\mathbf{X}} & E(\mathbf{X}\mathbf{X}') \end{pmatrix} + (n-1)\rho_y \begin{pmatrix} 1 & \mu_{\mathbf{X}}' \\ \mu_{\mathbf{X}} & E(\mathbf{X}_1\mathbf{X}_2') \end{pmatrix} \right] \qquad (3.4.4.7)$$

which we shall write as

$$T + (n-1)\rho_y W.$$

Next we consider the limiting value of the derivatives of $-\mathbf{S}_k(\hat{\alpha}_u, \hat{\boldsymbol{\beta}}_u)/k$. First we have

$$-\frac{\partial \mathbf{S}_k}{\partial \alpha} = \sum_{i=1}^{k} \sum_{j=1}^{n} \frac{\partial p_{ij}}{\partial \alpha}$$

where

$$\frac{\partial p_{ij}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[\frac{\exp(\alpha + \boldsymbol{\beta}'\mathbf{x})}{1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x})} \right]$$

$$= \frac{[1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x})]\exp(\alpha + \boldsymbol{\beta}'\mathbf{x}) - [\exp(\alpha + \boldsymbol{\beta}'\mathbf{x})]^2}{[1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{x})]^2}$$

$$= pq$$

where the ij subscript has been suppressed for clarity. Similarly $\partial p/\partial \boldsymbol{\beta}$ becomes $\mathbf{x}pq$, so that we may write $-\mathbf{S}_k'/k$ as the matrix U_k, that is, as

$$\frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{n} \begin{bmatrix} p_{ij}q_{ij} & \mathbf{x}_{ij}' p_{ij}q_{ij} \\ \mathbf{x}_{ij}p_{ij}q_{ij} & \mathbf{x}_{ij}\mathbf{x}_{ij}' p_{ij}q_{ij} \end{bmatrix}$$

where

$$p_{ij} = \Pr(Y_{ij} = 1)$$

under the usual logistic model and

$$q_{ij} = 1 - p_{ij}$$

and $\mathbf{x}_{ij}$ is the $(p \times 1)$ vector of observations on unit j in cluster i.

We wish to determine the limiting value of U_k as k gets large. As shown previously, as

$$k \longrightarrow \infty,$$

then

$$\hat{\alpha}_u \xrightarrow{p} \alpha_0 \ (= \text{logit } F)$$

and

$$\hat{\boldsymbol{\beta}}_u \xrightarrow{p} \boldsymbol{\beta}_0 \ (=0),$$

so that, for each i and j,

$$p_{ij} \xrightarrow{p} F$$

and

$$q_{ij} \xrightarrow{p} G \ (=1-F)$$

where

$$F = \Pr(Y=1 \mid \boldsymbol{\beta}=\mathbf{0})$$

under the correlated logistic model. Finally

$$U_k \xrightarrow{p} U$$

where

$$U = \begin{bmatrix} nFG & nFG\mu_{\mathbf{X}}' \\ nFG\mu_{\mathbf{X}} & nFG\mathbf{XX}' \end{bmatrix} = n\,\sigma_y^2 \begin{bmatrix} 1 & \mu_{\mathbf{X}}' \\ \mu_{\mathbf{X}} & \mathbf{XX}' \end{bmatrix}$$

Thus the asymptotic distribution of the usual estimator (when $\boldsymbol{\beta}$ is zero) is given by

$$\sqrt{k}\begin{pmatrix}\hat{\alpha}_u - \alpha_0 \\ \hat{\boldsymbol{\beta}}_u - \mathbf{0}\end{pmatrix} \sim MVN(0, U^{-1}VU^{-1})$$

where $U^{-1}VU^{-1}$ can be written

$$T^{-1} + (n-1)\rho_y T^{-1}WT^{-1}$$

where T and W have been defined above.

Moreover, the misspecificaion factor is

$$I + (n-1)\rho_y T^{-1}W$$

an expression quite analogous to that given in the multiple covariate case for the continuous dependent variable by Scott and Holt(1982, p850).

3.4.5 Dummy variables estimator

The dummmy variables estimator for the regression coefficients, denoted by $\hat{\boldsymbol{\beta}}_d$, was defined in section 3.3.4, where the following estimating equations were derived (see expressions (3.3.4.1) and (3.3.4.2))

$$\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ijm}(z_{ij} - p_{ij}), \; m=1,\ldots p$$

$$\sum_{j=1}^{n}(z_{ij} - p_{ij}), \; i=1,\ldots k \qquad (3.4.5.1)$$

where p_{ij} is the probability that Z_{ij} is 1 under the dummy varibles model, that is,

$$p_{ij} = \Pr(Z_{ij}=1|\mathbf{x}_{ij},\mathbf{v}_i)$$

and $\mathbf{v}_i$ is the indicator vector for cluster i. The expected value of these equations was considered in Section 3.3.4, where caution was exercised because the number of nuisance parameters, the γ's, approaches infinity as k does. In particular, we did not use quite the same approach to deriving these expectations as were used for the usual estimator. Similarly, in this section, we use a slightly different approach from that employed in the preceding section for the usual estimator in order to obtain both the asymptotic variance and misspecification factor for the dummy variables estimator. Again, as with the usual estimator, because

$$\boldsymbol{\beta} = \mathbf{0}$$

is a sufficient condition for the consistency of the dummy variables estimator, we concentrate on examining the variance and misspecifiation only when $\boldsymbol{\beta}$ is $\mathbf{0}$.

The assumption is made in this chapter that the expected value of the covariates is zero. This assumption simplifies the following calculations, but is not necessary for the general conclusions to be drawn in this chapter and in this thesis.

First we consider the expectation of the estimating equations at the limiting value of the dummy varibles estimator. If expression (3.4.5.1) defines the estimating equations $\mathbf{S}_k(\boldsymbol{\beta},\boldsymbol{\gamma})$, then $\mathbf{S}_k(\boldsymbol{\beta},\boldsymbol{\gamma})$ can be considered as the sum of k components which are independently distributed, namely,

$$\sum_{j=1}^{n} \mathbf{x}_j (z_j - p_j)$$

$$\sum_{j=1}^{n} (z_j - p_j)$$

Taking the expectation with respect to the correlated logistic distribution, and assuming that the marginal distribution of (X,Z) is the same for each unit, We get

$$n\, E[\mathbf{X}(Z-P)]$$

$$n\, E(Z-P)$$

In section 3.3.4, it was shown that, if

$$\boldsymbol{\beta} = \mathbf{0},$$

then

$$E(Z) = F$$

where F is the probability that Z is 1 under the correlated logistic model.

It was also shown in section 3.3.2 that

$$\boldsymbol{\beta} = \mathbf{0}$$

implies that

$$\boldsymbol{\beta}_0 = \mathbf{0}$$

where $\boldsymbol{\beta}_0$ is the limiting value of $\hat{\boldsymbol{\beta}}_d$, the dummy variables estimator of $\boldsymbol{\beta}$, the vector of regression coefficients. In this limiting condition, P can take only n-1 values, namely

$$p_r = r/n, \quad r=1,\ldots n-1.$$

For a further explanation, in particular, why P cannot take on values 0 or 1, see Appendix B.3. Now $Pr(P=p_r)$ can be written as s_r where

s_r is the probability that r of the Y's are 1, adjusted for the fact that r cannot be 0 or 1. Moreover, s_r can be written in terms of expressions developed in section 3.4.3, namely,

$$s_r = g_r\, c$$

or

$$s_r = {}_nC_r\, h_r\, c$$

where g_r is the probability that r of the Y's are 1, the remaining (n-r) Y's being 0, and h_r is the probability that a particular set of r Y's are 1, the rest being 0, and

$$1/c = 1-g_0-g_n = 1-h_0-h_n .$$

In paritcular, for the correlated logistic model,

$$h_r = \exp(\alpha_r) \Big/ \sum_{h=0}^{n} \exp(\alpha_h)$$

The expectation of P may be calculated as follows

$$E(P) = \sum_{r=1}^{n-1} p_r\, s_r$$

$$= \sum_{r=1}^{n-1} \frac{r}{n}\, {}_nC_r\, h_r\, c$$

$$= \sum_{r=1}^{n-1} {}_{n-1}C_{r-1}\, h_r\, c$$

Now h_r may be interpreted as the probability that Y_1 and exactly (r-1) of the other Y's are 1. Premultiplication of h_r by ${}_{n-1}C_{r-1}$ allows one to chose all other sets of (r-1) Y's. In other words

$${}_{n-1}C_{r-1}\, h_r = \Pr(\text{a specfic Y is 1})$$

$$= f_1$$

which is the marginal probability that Y is 1, conditional on the fact that we ignore clusters with all Y's equal to 1 or 0. Hence we have shown that

$$E(P) = E(Z)$$

for all clusters, where Z is the random varible Y with no possibility that Y is 0 or 1. The distribution of Z is simply

$$\Pr(Z = j) = \Pr(Y = j)\, c, \quad j=0,1,$$

where c was defined above. The joint distribution of the random vector Z may also be defined as

$$\Pr(\mathbf{Z}=\mathbf{z}) = \Pr(\mathbf{Y}=\mathbf{y})\, c.$$

This is further discussed in Apppendix B.2. Thus our estimating equations may be written

$$kn\ E[\mathbf{X}(Z-P)]$$

and

$$kn\ E(Z-P).$$

We have shown above that

$$E(Z-P) = 0$$

Further

$$\begin{aligned} E[\mathbf{X}(Z-P)] &= E_X[\ \mathbf{X}\ E_{Z|X}\ (Z-P)] \\ &= E_X(\mathbf{X} \times 0) \\ &= \mathbf{0} \end{aligned}$$

so that the expected values of the estimating equations values at the limiting values of the estimators is zero.

Table 4.11

Limiting value of the usual estimator

Clusters of size 10

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050	0.000	0.664	1.559	2.244
	0.200	0.000	0.593	1.473	2.162
	0.500	0.000	0.483	1.405	2.110
	0.800	0.000	0.340	1.357	2.088
0.00					
	0.050	0.000	0.693	1.607	2.299
	0.200	0.000	0.689	1.597	2.287
	0.500	0.000	0.680	1.587	2.280
	0.800	0.000	0.669	1.576	2.269
0.10					
	0.050	0.000	0.721	1.657	2.355
	0.200	0.000	0.786	1.730	2.426
	0.500	0.000	0.882	1.820	2.509
	0.800	0.000	0.999	1.947	2.618
0.20					
	0.050	0.000	0.750	1.706	2.412
	0.200	0.000	0.883	1.868	2.573
	0.500	0.000	1.084	2.071	2.767
	0.800	0.000	1.314	2.335	3.018
0.50					
	0.050	0.000	0.835	1.858	2.586
	0.200	0.000	1.177	2.297	3.038
	0.500	0.000	1.690	2.846	3.582
	0.800	0.000	2.218	3.424	4.154
1.00					
	0.050	0.000	0.979	2.119	2.890
	0.200	0.000	1.687	3.076	3.893
	0.500	0.000	2.795	4.297	5.125
	0.800	0.000	4.069	5.637	6.466

$$kn\ \sigma_z^2\ [E(\mathbf{XX}') + (n-1)\ \rho_z\ E(\mathbf{X}_1\mathbf{X}_2')] .$$

The variance-covariance matrix for the k equations corresponding to the k dummy variables is calculated next. The discussion assumes that the number of dummy variables parameters becomes infinitely large, but, if the sample space of the covariates has a finite number of elements, it may be argued that there are at most a finite number of unique dummy variables parameters (for a simple example, see Appendix C.3). However, this finiteness does not affect the results given in the rest of this chapter. The variance for each equation is

$$\mathrm{Var}\ [\sum_{j=1}^{n} (Z_j - P_j)]$$

Since

$$E(Z_j) = E(P_j) = F$$

this becomes

$$\mathrm{Var}(\sum_{j=1}^{n} Z_j) = n\ \sigma_z^2\ (1 + \rho_z).$$

Since the clusters are independent the off-diagonal elements of this dispersion matrix are 0.

Finally we examine the covariance matrix of the two sets of estimating equations namely, the matrix whose (h,m)'th element is

$$\mathrm{cov}[\sum_{i=1}^{k} \sum_{j=1}^{n} X_{ijh}(Z_{ij}-P_{ij}),\ \sum_{j=1}^{n} (Z_{mj}-P_{mj})]$$

which, because the observations are independent between clusters, becomes

$$\mathrm{cov}\Big[\sum_{j=1}^{n} X_{mjh}(Z_{mj}-P_{mj}),\ \sum_{j=1}^{n} (Z_{mj}-P_{mj})\Big]$$

and, since

$$E(Z_{mj}) = E(P_{mj}),$$

this becomes

$$\mathrm{cov}\Big[\sum_{j=1}^{n} (X_{mjh}\ Z_{mj}),\ \sum_{j=1}^{n} Z_{mj}\Big].$$

Now, because the marginal and joint distributions are the same for each j,

$$j=1,\ldots,n$$

this may be written

$$n\ \mathrm{cov}(X_h Z,Z) + n(n-1)\ \mathrm{cov}(X_h Z_1,Z_2)$$

$$= n\ \sigma_z^2\ [1 + (n-1)\rho_z]\ E(X_h).$$

Thus we have determined the dispersion matrix of $\mathbf{S}_k(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)$. Standardizing by dividing $\mathbf{S}_k(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)$ by the square root of k, we have the result

$$\frac{1}{\sqrt{k}}\ \mathbf{S}_k(\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0) \sim (\mathbf{0},V)$$

where V can be partitioned into four submatrices, namely

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{bmatrix},$$

where

$$V_{11} = n\ \sigma_z^2\ [E(\mathbf{X}\mathbf{X}') + (n-1)\rho_z\ E(\mathbf{X}_1\mathbf{X}_2')]$$

which may be written as

$$n\ \sigma_z^2\ [T_d + (n-1)\rho_z\ W_d],$$

whereas V_{12} and V_{21} are zero because we have assumed that

$$E(\mathbf{X}) = \mathbf{0}.$$

The next step is to determine the limiting form of $-S_k''(\boldsymbol{\beta},\boldsymbol{\gamma})/k$ as k goes to infinity. The derivatives of the expressions (3.4.5.1) yields a matrix U_k of the form

$$\frac{1}{k}\sum_{j=1}^{n}\begin{pmatrix} \sum_{i=1}^{k} \mathbf{x}_{ij}\mathbf{x}_{ij}' p_{ij}q_{ij} & \cdots & & \\ \mathbf{x}_{1j}' p_{1j}q_{1j} & p_{1j}q_{1j} & 0 & \cdots \\ \mathbf{x}_{2j}' p_{2j}q_{2j} & 0 & p_{2j}q_{2j} & \cdots \\ \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \end{pmatrix} \qquad (3.4.5.7)$$

To invert this matrix we use the following result (see, for example, Graybill(1976)): for a matrix M which has been partitioned into four submatrices, such that

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

the inverse matrix M^{-1} can be partitioned in asimilar way, namely,

$$M^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}$$

We are interested in the submatrix U_k^{11}, which, from (3.4.5.7), can be written

$$\{\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n}\mathbf{x}_{ij}\mathbf{x}_{ij}'p_{ij}q_{ij}$$

$$-\frac{1}{k}\sum_{i=1}^{k}[(\sum_{j=1}^{n}\mathbf{x}_{ij}p_{ij}q_{ij})(\sum_{j=1}^{n}\mathbf{x}_{ij}'p_{ij}q_{ij})/(\sum_{j=1}^{n}p_{ij}q_{ij})]\}^{-1}$$

When

$$\boldsymbol{\beta} = \mathbf{0},$$

as

$$\kappa \to \infty,$$

then

$$p_{ij} \xrightarrow{p} f_i,$$

$$q_{ij} \xrightarrow{p} g_i,$$

$$\sum_{j=1}^{n} p_{ij}q_{ij} \xrightarrow{p} nf_i g_i,$$

$$\sum_{j=1}^{n} \mathbf{x}_{ij}p_{ij}q_{ij} \xrightarrow{p} f_i g_i \sum_{j=1} \mathbf{x}_{ij},$$

and

$$\sum_{j=1}^{n} \mathbf{x}_{ij}'p_{ij}q_{ij} \xrightarrow{p} f_i g_i \sum_{j=1}^{n} \mathbf{x}_{ij}'$$

so that

$$(\sum_{j=1}^{n}\mathbf{x}_{ij}p_{ij}q_{ij})(\sum_{j=1}^{n}\mathbf{x}_{ij}'p_{ij}q_{ij})/(\sum_{j=1}^{n}p_{ij}q_{ij})$$

$$\xrightarrow{p}(\frac{1}{n} f_i g_i \sum_{j=1}^{n}\mathbf{x}_{ij}\sum_{j=1}^{n}\mathbf{x}_{ij}')$$

Now assuming that a law of large numbers holds so that, in general

$$\lim_{k \to \infty} \frac{1}{k}[f_k(\cdot)] = E[f(\cdot)]$$

then

$$U^{11} = \{E[f_i g_i(\sum_{j=1}^{n} \mathbf{X}_j \mathbf{X}_j')] - \frac{1}{n} E[f_i g_i(\sum_{j=1}^{n} \mathbf{X}_j \sum_{j=1}^{n} \mathbf{X}_j')]\}^{-1}$$

$$= (\sigma_z^2 \{nE(\mathbf{XX}') - \frac{1}{n}[nE(\mathbf{XX}')+n(n-1)E(\mathbf{X}_1\mathbf{X}_2')]\})^{-1}$$

$$= \{\sigma_z^2 (n-1) [E(\mathbf{XX}')-E(\mathbf{X}_1\mathbf{X}_2')]\}^{-1}$$

$$= [\sigma_z^2 (n-1) (T_d - W_d)]^{-1}$$

where T_d and W_d were defined above, and are similar to the matrices T and W defined for the usual estimator.

By the methodology established in Appendix B.4,

$$\sqrt{k}\, \hat{\boldsymbol{\theta}}_d \xrightarrow{D} (\mathbf{0}, U^{-1}VU^{-1})$$

where

$$\hat{\boldsymbol{\theta}}_d = (\hat{\boldsymbol{\beta}}_d, \hat{\boldsymbol{\gamma}}_d)'$$

but we are only interested in the estimation and testing of $\hat{\boldsymbol{\beta}}_d$. Since V has only two non-zero submatrices, V_{11} and V_{22}, the matrix of interest, namely, that in the upper left corner of $U^{-1}VU^{-1}$, can be written as

$$U^{11}V_{11}U^{11} + U^{12}V_{22}U^{21}$$

where U^{12} can be written as

$$- U^{11} U_{12} U_{22}^{-1}$$

but, as can be seen by examining the elements of (3.4.5.7), U_{12} has a limiting value of zero, so that the variance term of interest can be written as

$$U^{11} V_{11} U^{11}.$$

Hence we have the result

$$\sqrt{k}\, \hat{\beta}_d \xrightarrow{D} (0, U^{11} V_{11} U^{11}),$$

which shows that the asymptotic variance of $\hat{\beta}_d$ is

$$\frac{n}{k(n-1)^2 \sigma_z^2} [T_d - W_d]^{-1} [T_d + (n-1)\rho_z W_d] [T_d - W_d]^{-1}. \quad (3.4.5.8)$$

This expression gives a 'conditional' variance, conditional on the assumption that clusters with all Y values the same have been excluded (see Appendix B.1). For comparative purposes, in particular, to order to compare $\hat{\beta}_d$ with other estimators that do not exclude these clusters we must calculate an 'unconditional' asymptotic variance. It is shown in Appendix B.5 that the appropriate asymptotic variance is expression (3.4.5.8) multiplied by c, where c was defined in earlier in this section (and also in Appendix B.2) as

$$1/[1 - \Pr(\text{all Y's are 0}) - \Pr(\text{all Y's are 1})].$$

The misspecification factor for the variance of the dummy variables estimator is given by the following argument. In testing the hypothesis:

$$\beta = 0,$$

when one has assumed the usual logistic regression model, one uses the estimator $\hat{\theta}_d$ with an estimated variance given by the expression

$$[\neg S_N'(\hat{\theta}_d)]^{-1}$$

where

$$N = nk,$$

all clusters with equal Y's having been removed. This term $S_N''(\hat{\theta}_d)$ is the matrix of second derivatives of the usual logistic model evaluated at the estimate $\hat{\theta}_d$.

We now consider the asymptotic behaviour of $[-S_N''(\theta)]^{-1}$ under the correlated logistic model. It is easily seen that $S_N''(.)$ is equivalent to $S_k''(.)$ described earlier in this section. Since

$$-\frac{1}{k} S_k''(\hat{\theta}_d) \xrightarrow{p} U$$

we know that

$$\left[-\frac{1}{k} S_N''(\hat{\theta}_d) \right]^{-1} \xrightarrow{p} U^{-1}$$

that is, $[-S_N''(\hat{\theta}_d)]^{-1}$, as used in the test statistic, actually estimates U^{-1}/k. However the true variance of $\hat{\beta}_d$ is given by expression (3.4.5.8) multiplied by c, so that the misspecification factor in using $[-S_N''(\hat{\theta}_d)]^{-1}$ is

$$\frac{n}{n-1} U^{11} [T_d + (n-1)\rho_z W_d]c$$

which may be written as

$$\frac{nc}{n-1} [E(\mathbf{X}\mathbf{X}')-E(\mathbf{X}_1\mathbf{X}_2')]^{-1}$$

$$\times [E(\mathbf{X}\mathbf{X}')+(n-1)\rho_z E(\mathbf{X}_1\mathbf{X}_2')]. \qquad (3.4.5.9)$$

3.4.6 Summary

The variances obtained in the last five sections are summarized in Table 3.1. Recall that these expressions hold only when

$$\boldsymbol{\beta} = \mathbf{0}. \quad (3.4.6.1)$$

Also, for all estimators except the conditional it is assumed that

$$E(\mathbf{X}_1) = E(\mathbf{X}_2) = E(\mathbf{X}) = \mathbf{0} \quad (3.4.6.2)$$

This enables us to invert matrices but to ignore all other submatrices except the one presented in the table. Hence for all the entries in the table, we may replace $E(\mathbf{X}\mathbf{X}')$ with $\mathrm{Var}(\mathbf{X})$ and $E(\mathbf{X}_1\mathbf{X}_2')$ with $\mathrm{Cov}(\mathbf{X}_1,\mathbf{X}_2)$. It is conjectured that, at least when (3.4.6.1) holds, assumption (3.4.6.2) may not be necessary for the replacement of non-central second moments with central second moments.

Recall that the random vector $\mathbf{X}$ indicates that observations are taken on the same unit, and the vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ indicate that observations are taken on any two units within the same cluster.

Table 3.1

Asymptotic variance of four estimators

($\boldsymbol{\beta} = 0$)

Correlated logistic

$$\{n\, \sigma_y^2\, [E(\mathbf{XX}') + (n-1)\rho_y E(\mathbf{X}_1\mathbf{X}_2')]\}^{-1}$$

Conditional logistic

$$\{(n-1)\, \sigma_y^2\, (1-\rho_y)\, [E(\mathbf{XX}') - E(\mathbf{X}_1\mathbf{X}_2')]\}^{-1}$$

Usual logistic

$$U^{-1}\, n\, \sigma_y^2\, [E(\mathbf{XX}') + (n-1)\rho_y E(\mathbf{X}_1\mathbf{X}_2')]\, U^{-1}$$

where

$$U = n\, \sigma_y^2\, E(\mathbf{XX}')$$

Dummy varibles logistic

$$U_d^{-1}\, c\, n\, \sigma_z^2\, [E(\mathbf{XX}') + (n-1)\rho_z E(\mathbf{X}_1\mathbf{X}_2')]\, U_d^{-1}$$

where

$$U_d = (n-1)\, \sigma_z^2\, [E(\mathbf{XX}') - E(\mathbf{X}_1\mathbf{X}_2')]$$

and

$$c = 1/[1 - Pr(\text{all } Y\text{'s are } 0) - Pr(\text{all } Y\text{'s are } 1)]$$

CHAPTER 4

Asymptotic results in particular cases

4.1 Introduction

In this chapter we examine the bias, variance (relative efficiency) and variance misspecification factor of the estimators, for one or two covariates, and varying number of units per cluster.

It is shown that, in the case of a single covariate,

1. the usual estimator, $\hat{\beta}_u$, is in general biased for all cluster sizes, but, under certain conditions, it may be a more efficient estimator than $\hat{\beta}_R$, the maximum likelihood estimator (for Rosner's constrained version of the correlated logistic regression model).

2. the dummy variables estimator, $\hat{\beta}_d$, is biased but the amount of bias decreases with cluster size. However, the relative efficiency also decreases with cluster size unless both intraclass coefficients are very close to zero.

3. the conditional estimator, $\hat{\beta}_c$, has low relative efficiency for small cluster size, and, although the efficiency increases with cluster size, it attains a limit which is less than 1, except under special conditions.

The relative efficiency of $\hat{\beta}_c$ is investigated for non-zero values of β, and it is shown that the efficiency decreases with increasing values of $|\beta|$.

We also investigate the behaviour of the usual estimator for two covariates, and discover that, for

$$\beta_1 = 0, \ \beta_2 \neq 0,$$

1. the usual estimator of β_2 behaves as described for the single covariate case

2. the usual estimator of β_1 is always negatively biased.

In this chapter is is assumed that k, the number of clusters, is very large, that is,

$$k \to \infty.$$

Thus we are dealing with bias and variance that are asymptotic in k but calculated for finite cluster size. Since we are interested in bias, variance, etc. as a function of n, the cluster size, we shall use the term bias and variance although these results are derived asymptotically in k, the number of clusters.

Also, since we are interested in measuring the effect of a single intraclass (intracluster) correlation coefficient, namely, ρ_y, on the bias and variance, we examine the behaviour of estimators only under Rosner's version of the correlated logisitc model.

4.2 Bias of the usual estimator

It was shown in section 3.3.3 that the usual estimator of the regression coefficient is, in general, biased, and that a sufficient condition for unbiasedness is that

$$\beta = 0,$$

that is, that there is no covariate effect on the dependent variable. However, it may be that the bias is small, or becomes small for large cluster size. Also there may be other conditions under which the estimator is unbiased. In this section we investigate the bias of the usual estimator in two cases

1. a single binary covariate with increasing cluster size.

2. two correlated binary covariates for cluster size two

4.2.1 Single binary covariate

When the single covariate is binary, and the intracluster correlation is ignored, the data may be written as a single contingency table, with the binary covariate defining the rows and the response variable defining the column. The following indicates the form of a typical contingency table.

		dependent variable 0	dependent variable 1	
covariate	0	n_{00}	n_{01}	n_{0+}
	1	n_{10}	n_{11}	n_{1+}
		n_{+0}	n_{+1}	N

where n_{ij} is the number of observations with covariate value i and response value j, and + indicated summation of over all possible values.

It is well known that, under a binomial error assumption, the maximum likelihood estimator of the odds ratio is

$$(n_{00}\, n_{11})/(n_{01}\, n_{10}). \qquad (4.2.1.1)$$

Futhermore, if it is assumed that distribution of the dependent variable follows the logistic model, then it can be shown (see Appendix C.1) that the maximum likelihood estimator of β, the coefficient of the covariate in the usual logistic model, is

$$\hat{\beta}_u = \log\,(n_{00}\, n_{11}/n_{01}\, n_{10}),$$

that is, $\hat{\beta}_u$ is the maximum likelihood estimator of the log odds ratio, or, in other words, the maximum likelihood estimator of the odds ratio is $\exp(\hat{\beta}_u)$.

If we now assume the correlated logistic model, we regard the N observations as k clusters of size n. It therefore follows that the frequency n_{ij} may be considered as the observed value of a random variable N_{ij} which is the sum of k independent random variables R_{ij}, where R_{ij} can assume values $0,\ldots,n$. Hence the probability distribution of R_{ij} may be written

$$\Pr(R_{ij}=r) = \Pr(\text{exactly } r \text{ of the } n \text{ units have } x=i \text{ and } y=j)$$
$$= \sum_{\substack{r\ (x,y)'s \\ =(i,j)}} f(y|x)\ p(x)$$

where $f(y|x)$ is the conditional distribution of Y and $p(x)$ is the marginal distribution of X.

In order to evaluate the distribution of R_{ij} we shall assume that $f(y|x)$ is Rosner's version of the correlated logistic with parameters n, 0.5, $\rho_{y|x}$ and β, and that $p(x)$ is the correlated binary distribution with parameters n, 0.5 and ρ_x (see section 1.4.3).

For the odds ratio estimator (4.2.1.1) each random variable N_{ij} is the sum of k independent R_{ij}'s, so that, under some regularity conditions on the joint distribution $f(y,x)$, as

$$k \to \infty,$$

$$\frac{n_{00}\ n_{11}}{n_{10}\ n_{01}} \xrightarrow{p} \frac{E(N_{00})\ E(N_{11})}{E(N_{10})\ E(N_{01})}$$

$$= \frac{E(R_{00})\ E(R_{11})}{E(R_{10})\ E(R_{01})}. \qquad (4.2.1.2)$$

For fixed n, an algebraic expression may be obtained for (4.2.1.2), but, even for

$$n = 2$$

this expression is very complicated (see Appendix C.2). However, we may numerically evaluate the expression for varying values of ρ_x, $\rho_{y|x}$, and β. The program BIASU (see Appendix D.1) was written to perform such calculations.

Tables 4.1 to 4.25 give the values of

1. the limiting value of the usual estimator, $\hat{\beta}_u$, denoted by β_0,

2. the bias, $\beta_0-\beta$,

3. the relative bias, $(\beta-\beta_0)/\beta$,

4. the limiting value of the odds ratio estimator based on the usual estimator, $\exp(\beta_0)$

5. the inflation factor for the odds ratio estimator, $\exp(\beta_0)/\exp(\beta)$

for cluster sizes

$$n = 2,5,10,25,50$$

and for varying values of ρ_x, $\rho_{y|x}$ and β.

The inflation factor is the amount of error in $\exp(\beta_0)$ as an estimator of $\exp(\beta)$, that is, the amount by which $\exp(\beta_0)$ must be multiplied to yield $\exp(\beta)$. Values of the inflation factor less than 1 mean that $\hat{\beta}_u$ is negatively biased, and values greater than 1 mean that $\hat{\beta}_u$ is positively biased.

These tables reveal the following properties of the bias of the usual estimator:

1. it increases with increasing values of $\rho_{y|x}$.

2. it is negative for negative and small positive values of ρ_x.

3. it is positive for moderate to large values of ρ_x.

4. it generally increases with the absolute value of β, but, for values of ρ_x near 0.0, it may not do so.

5. it slowly increases with cluster size, so that at cluster size of 50, say, and

$$\exp(\beta) = 2.0,\ \rho_x = 0.2,\ \rho_{y|x} = 0.2,$$

the relative bias is 64 percent.

Table 4.1

Limiting value of the usual estimator

Clusters of size 2

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-1.00					
	0.050	0.000	0.660	1.543	2.221
	0.200	0.000	0.560	1.341	1.972
	0.500	0.000	0.357	0.916	1.434
	0.800	0.000	0.147	0.420	0.736
-0.50					
	0.050	0.000	0.676	1.576	2.261
	0.200	0.000	0.626	1.471	2.130
	0.500	0.000	0.522	1.242	1.832
	0.800	0.000	0.413	0.973	1.463
0.00					
	0.050	0.000	0.693	1.609	2.302
	0.200	0.000	0.692	1.604	2.292
	0.500	0.000	0.689	1.575	2.242
	0.800	0.000	0.682	1.524	2.151
0.10					
	0.050	0.000	0.696	1.616	2.310
	0.200	0.000	0.706	1.631	2.325
	0.500	0.000	0.722	1.644	2.326
	0.800	0.000	0.736	1.637	2.293
0.50					
	0.050	0.000	0.710	1.643	2.343
	0.200	0.000	0.759	1.739	2.459
	0.500	0.000	0.858	1.925	2.679
	0.800	0.000	0.956	2.112	2.899
1.00					
	0.050	0.000	0.726	1.676	2.384
	0.200	0.000	0.827	1.877	2.632
	0.500	0.000	1.029	2.302	3.171
	0.800	0.000	1.239	2.797	3.867

Table 4.2

Asymptotic bias of the usual estimator.

Clusters of size 2

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-1.00					
	0.050	0.000	-0.033	-0.067	-0.082
	0.200	0.000	-0.134	-0.268	-0.330
	0.500	0.000	-0.336	-0.693	-0.869
	0.800	0.000	-0.547	-1.190	-1.566
-0.50					
	0.050	0.000	-0.017	-0.034	-0.041
	0.200	0.000	-0.067	-0.138	-0.172
	0.500	0.000	-0.171	-0.368	-0.470
	0.800	0.000	-0.280	-0.636	-0.840
0.00					
	0.050	0.000	0.000	0.000	-0.001
	0.200	0.000	-0.001	-0.006	-0.010
	0.500	0.000	-0.005	-0.034	-0.061
	0.800	0.000	-0.012	-0.086	-0.151
0.10					
	0.050	0.000	0.003	0.006	0.008
	0.200	0.000	0.013	0.021	0.023
	0.500	0.000	0.029	0.034	0.024
	0.800	0.000	0.043	0.027	-0.010
0.50					
	0.050	0.000	0.017	0.033	0.040
	0.200	0.000	0.066	0.130	0.157
	0.500	0.000	0.164	0.316	0.377
	0.800	0.000	0.263	0.502	0.597
1.00					
	0.050	0.000	0.033	0.067	0.082
	0.200	0.000	0.133	0.268	0.330
	0.500	0.000	0.336	0.692	0.868
	0.800	0.000	0.546	1.188	1.564

Table 4.3

Relative bias of the usual estimator

Clusters of size 2

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
		Values of β (followed by values of exp(β))			
-1.00					
	0.050		-0.048	-0.041	-0.036
	0.200		-0.193	-0.167	-0.143
	0.500		-0.485	-0.431	-0.377
	0.800		-0.788	-0.739	-0.680
-0.50					
	0.050		-0.024	-0.021	-0.018
	0.200		-0.097	-0.086	-0.075
	0.500		-0.247	-0.228	-0.204
	0.800		-0.404	-0.395	-0.365
0.00					
	0.050		0.000	0.000	0.000
	0.200		-0.001	-0.004	-0.004
	0.500		-0.007	-0.021	-0.026
	0.800		-0.017	-0.053	-0.066
0.10					
	0.050		0.005	0.004	0.003
	0.200		0.018	0.013	0.010
	0.500		0.042	0.021	0.010
	0.800		0.062	0.017	-0.004
0.50					
	0.050		0.024	0.021	0.018
	0.200		0.095	0.081	0.068
	0.500		0.237	0.196	0.164
	0.800		0.379	0.312	0.259
1.00					
	0.050		0.048	0.041	0.036
	0.200		0.192	0.167	0.143
	0.500		0.485	0.430	0.377
	0.800		0.788	0.738	0.679

Table 4.4

Limiting value of estimator of odds ratio

Clusters of size 2

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-1.00					
	0.050	1.000	1.934	4.677	9.214
	0.200	1.000	1.750	3.824	7.187
	0.500	1.000	1.429	2.500	4.194
	0.800	1.000	1.158	1.522	2.088
-0.50					
	0.050	1.000	1.967	4.835	9.594
	0.200	1.000	1.870	4.355	8.417
	0.500	1.000	1.685	3.462	6.249
	0.800	1.000	1.511	2.647	4.317
0.00					
	0.050	1.000	2.000	4.998	9.993
	0.200	1.000	1.999	4.972	9.897
	0.500	1.000	1.991	4.832	9.410
	0.800	1.000	1.977	4.589	8.596
0.10					
	0.050	1.000	2.007	5.032	10.075
	0.200	1.000	2.025	5.107	10.229
	0.500	1.000	2.059	5.174	10.241
	0.800	1.000	2.087	5.139	9.903
0.50					
	0.050	1.000	2.034	5.168	10.412
	0.200	1.000	2.137	5.692	11.697
	0.500	1.000	2.358	6.857	14.577
	0.800	1.000	2.601	8.264	18.163
1.00					
	0.050	1.000	2.068	5.344	10.852
	0.200	1.000	2.285	6.537	13.908
	0.500	1.000	2.799	9.992	23.821
	0.800	1.000	3.453	16.403	47.779

Table 4.5

Inflation factor for estimator of odds ratio

Clusters of size 2

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-1.00					
	0.050	1.000	0.967	0.935	0.921
	0.200	1.000	0.875	0.765	0.719
	0.500	1.000	0.714	0.500	0.419
	0.800	1.000	0.579	0.304	0.209
-0.50					
	0.050	1.000	0.983	0.967	0.959
	0.200	1.000	0.935	0.871	0.842
	0.500	1.000	0.843	0.692	0.625
	0.800	1.000	0.756	0.529	0.432
0.00					
	0.050	1.000	1.000	1.000	0.999
	0.200	1.000	0.999	0.994	0.990
	0.500	1.000	0.995	0.966	0.941
	0.800	1.000	0.989	0.918	0.860
0.10					
	0.050	1.000	1.003	1.006	1.008
	0.200	1.000	1.013	1.021	1.023
	0.500	1.000	1.029	1.035	1.024
	0.800	1.000	1.044	1.028	0.990
0.50					
	0.050	1.000	1.017	1.034	1.041
	0.200	1.000	1.068	1.138	1.170
	0.500	1.000	1.179	1.371	1.458
	0.800	1.000	1.301	1.653	1.816
1.00					
	0.050	1.000	1.034	1.069	1.085
	0.200	1.000	1.143	1.307	1.391
	0.500	1.000	1.400	1.998	2.382
	0.800	1.000	1.726	3.281	4.778

Table 4.6

Limiting value of the usual estimator

Clusters of size 5

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.25					
	0.050	0.000	0.660	1.547	2.228
	0.200	0.000	0.567	1.395	2.065
	0.500	0.000	0.388	1.155	1.864
	0.800	0.000	0.183	0.829	1.694
-0.10					
	0.050	0.000	0.680	1.584	2.271
	0.200	0.000	0.641	1.514	2.192
	0.500	0.000	0.564	1.385	2.054
	0.800	0.000	0.475	1.198	1.830
0.00					
	0.050	0.000	0.693	1.608	2.300
	0.200	0.000	0.691	1.596	2.283
	0.500	0.000	0.680	1.559	2.233
	0.800	0.000	0.664	1.497	2.139
0.10					
	0.050	0.000	0.706	1.633	2.330
	0.200	0.000	0.740	1.680	2.376
	0.500	0.000	0.797	1.736	2.423
	0.800	0.000	0.851	1.789	2.457
0.50					
	0.050	0.000	0.758	1.733	2.449
	0.200	0.000	0.940	2.023	2.764
	0.500	0.000	1.269	2.466	3.217
	0.800	0.000	1.605	2.921	3.675
1.00					
	0.050	0.000	0.824	1.860	2.601
	0.200	0.000	1.196	2.479	3.289
	0.500	0.000	1.900	3.501	4.360
	0.800	0.000	2.695	4.750	5.646

Table 4.7

Asymptotic bias of the usual estimator

Clusters of size 5

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.25					
	0.050	0.000	-0.033	-0.063	-0.074
	0.200	0.000	-0.126	-0.215	-0.237
	0.500	0.000	-0.305	-0.454	-0.439
	0.800	0.000	-0.510	-0.781	-0.609
-0.10					
	0.050	0.000	-0.013	-0.026	-0.031
	0.200	0.000	-0.052	-0.095	-0.111
	0.500	0.000	-0.129	-0.224	-0.248
	0.800	0.000	-0.218	-0.411	-0.473
0.00					
	0.050	0.000	0.000	-0.001	-0.002
	0.200	0.000	-0.002	-0.013	-0.020
	0.500	0.000	-0.013	-0.051	-0.069
	0.800	0.000	-0.029	-0.113	-0.164
0.10					
	0.050	0.000	0.013	0.023	0.027
	0.200	0.000	0.047	0.070	0.073
	0.500	0.000	0.104	0.127	0.121
	0.800	0.000	0.158	0.180	0.154
0.50					
	0.050	0.000	0.065	0.123	0.146
	0.200	0.000	0.247	0.413	0.462
	0.500	0.000	0.576	0.857	0.914
	0.800	0.000	0.912	1.312	1.373
1.00					
	0.050	0.000	0.131	0.250	0.299
	0.200	0.000	0.502	0.869	0.987
	0.500	0.000	1.206	1.892	2.057
	0.800	0.000	2.002	3.140	3.344

Table 4.8

Relative bias of the usual estimator

Clusters of size 5

Values of β (followed by values of exp(β)).

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.25					
	0.050		-0.047	-0.039	-0.032
	0.200		-0.182	-0.133	-0.103
	0.500		-0.441	-0.282	-0.191
	0.800		-0.736	-0.485	-0.264
-0.10					
	0.050		-0.019	-0.016	-0.014
	0.200		-0.075	-0.059	-0.048
	0.500		-0.186	-0.139	-0.108
	0.800		-0.315	-0.255	-0.205
0.00					
	0.050		0.000	-0.001	-0.001
	0.200		-0.004	-0.008	-0.009
	0.500		-0.018	-0.031	-0.030
	0.800		-0.042	-0.070	-0.071
0.10					
	0.050		0.019	0.015	0.012
	0.200		0.068	0.044	0.032
	0.500		0.150	0.079	0.052
	0.800		0.228	0.112	0.067
0.50					
	0.050		0.094	0.077	0.063
	0.200		0.357	0.257	0.201
	0.500		0.831	0.532	0.397
	0.800		1.316	0.815	0.596
1.00					
	0.050		0.189	0.156	0.130
	0.200		0.725	0.540	0.429
	0.500		1.740	1.175	0.893
	0.800		2.888	1.951	1.452

Table 4.9

Limiting value of estimator of odds ratio

Clusters of size 5

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.25					
	0.050	1.000	1.935	4.697	9.283
	0.200	1.000	1.763	4.034	7.889
	0.500	1.000	1.474	3.174	6.448
	0.800	1.000	1.201	2.290	5.441
-0.10					
	0.050	1.000	1.974	4.872	9.692
	0.200	1.000	1.899	4.545	8.953
	0.500	1.000	1.757	3.996	7.802
	0.800	1.000	1.608	3.314	6.23[illegible]
0.00					
	0.050	1.000	2.000	4.994	9.979
	0.200	1.000	1.995	4.934	9.805
	0.500	1.000	1.975	4.753	9.331
	0.800	1.000	1.943	4.467	8.490
0.10					
	0.050	1.000	2.026	5.119	10.275
	0.200	1.000	2.096	5.363	10.762
	0.500	1.000	2.219	5.676	11.281
	0.800	1.000	2.343	5.986	11.664
0.50					
	0.050	1.000	2.135	5.656	11.573
	0.200	1.000	2.561	7.559	15.870
	0.500	1.000	3.559	11.775	24.949
	0.800	1.000	4.978	18.561	39.458
1.00					
	0.050	1.000	2.280	6.422	13.479
	0.200	1.000	3.305	11.925	26.827
	0.500	1.000	6.683	33.156	78.220
	0.800	1.000	14.802	115.531	283.242

Table 4.10

Inflation factor for estimator of odds ratio

Clusters of size 5

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.25					
	0.050	1.000	0.968	0.939	0.928
	0.200	1.000	0.881	0.807	0.789
	0.500	1.000	0.737	0.635	0.645
	0.800	1.000	0.600	0.458	0.544
-0.10					
	0.050	1.000	0.987	0.974	0.969
	0.200	1.000	0.949	0.909	0.895
	0.500	1.000	0.879	0.799	0.780
	0.800	1.000	0.804	0.663	0.623
0.00					
	0.050	1.000	1.000	0.999	0.998
	0.200	1.000	0.998	0.987	0.980
	0.500	1.000	0.987	0.951	0.933
	0.800	1.000	0.972	0.893	0.849
0.10					
	0.050	1.000	1.013	1.024	1.028
	0.200	1.000	1.048	1.073	1.076
	0.500	1.000	1.109	1.135	1.128
	0.800	1.000	1.171	1.197	1.166
0.50					
	0.050	1.000	1.067	1.131	1.157
	0.200	1.000	1.280	1.512	1.587
	0.500	1.000	1.779	2.355	2.495
	0.800	1.000	2.489	3.712	3.946
1.00					
	0.050	1.000	1.140	1.284	1.348
	0.200	1.000	1.653	2.385	2.683
	0.500	1.000	3.341	6.631	7.822
	0.800	1.000	7.401	23.106	28.324

Table 4.11

Limiting value of the usual estimator

Clusters of size 10

Values of β (followed by values of $\exp(\beta)$)

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050	0.000	0.664	1.559	2.244
	0.200	0.000	0.593	1.473	2.162
	0.500	0.000	0.483	1.405	2.110
	0.800	0.000	0.340	1.357	2.088
0.00					
	0.050	0.000	0.693	1.607	2.299
	0.200	0.000	0.689	1.597	2.287
	0.500	0.000	0.680	1.587	2.280
	0.800	0.000	0.669	1.576	2.269
0.10					
	0.050	0.000	0.721	1.657	2.355
	0.200	0.000	0.786	1.730	2.426
	0.500	0.000	0.882	1.820	2.509
	0.800	0.000	0.999	1.947	2.618
0.20					
	0.050	0.000	0.750	1.706	2.412
	0.200	0.000	0.883	1.868	2.573
	0.500	0.000	1.084	2.071	2.767
	0.800	0.000	1.314	2.335	3.018
0.50					
	0.050	0.000	0.835	1.858	2.586
	0.200	0.000	1.177	2.297	3.038
	0.500	0.000	1.690	2.846	3.582
	0.800	0.000	2.218	3.424	4.154
1.00					
	0.050	0.000	0.979	2.119	2.890
	0.200	0.000	1.687	3.076	3.893
	0.500	0.000	2.795	4.297	5.125
	0.800	0.000	4.069	5.637	6.466

Table 4.12

Asymptotic bias of the usual estimator

Clusters of size 10

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050	0.000	-0.029	-0.051	-0.058
	0.200	0.000	-0.100	-0.136	-0.140
	0.500	0.000	-0.211	-0.205	-0.193
	0.800	0.000	-0.353	-0.253	-0.214
0.00					
	0.050	0.000	0.000	-0.002	-0.003
	0.200	0.000	-0.004	-0.012	-0.015
	0.500	0.000	-0.013	-0.022	-0.022
	0.800	0.000	-0.024	-0.034	-0.033
0.10					
	0.050	0.000	0.028	0.047	0.053
	0.200	0.000	0.093	0.121	0.124
	0.500	0.000	0.189	0.211	0.207
	0.800	0.000	0.305	0.337	0.315
0.20					
	0.050	0.000	0.056	0.097	0.109
	0.200	0.000	0.190	0.258	0.270
	0.500	0.000	0.391	0.462	0.465
	0.800	0.000	0.620	0.726	0.715
0.50					
	0.050	0.000	0.142	0.248	0.284
	0.200	0.000	0.484	0.688	0.735
	0.500	0.000	0.997	1.237	1.279
	0.800	0.000	1.525	1.814	1.852
1.00					
	0.050	0.000	0.286	0.509	0.588
	0.200	0.000	0.994	1.466	1.591
	0.500	0.000	2.102	2.688	2.822
	0.800	0.000	3.375	4.028	4.164

Table 4.13

Relative bias of the usual estimator

Clusters of size 10

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050		-0.041	-0.032	-0.025
	0.200		-0.145	-0.085	-0.061
	0.500		-0.304	-0.127	-0.084
	0.800		-0.509	-0.157	-0.093
0.00					
	0.050		-0.001	-0.001	-0.001
	0.200		-0.006	-0.008	-0.007
	0.500		-0.019	-0.014	-0.010
	0.800		-0.035	-0.021	-0.015
0.10					
	0.050		0.040	0.029	0.023
	0.200		0.134	0.075	0.054
	0.500		0.272	0.131	0.090
	0.800		0.441	0.210	0.137
0.20					
	0.050		0.081	0.060	0.048
	0.200		0.274	0.160	0.117
	0.500		0.564	0.287	0.202
	0.800		0.895	0.451	0.311
0.50					
	0.050		0.205	0.154	0.123
	0.200		0.698	0.428	0.319
	0.500		1.438	0.768	0.556
	0.800		2.200	1.127	0.804
1.00					
	0.050		0.412	0.316	0.255
	0.200		1.433	0.911	0.691
	0.500		3.033	1.670	1.226
	0.800		4.870	2.503	1.808

Table 4.14

Limiting value of estimator of odds ratio

Clusters of size 10

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050	1.000	1.943	4.752	9.435
	0.200	1.000	1.809	4.363	8.692
	0.500	1.000	1.620	4.075	8.247
	0.800	1.000	1.405	3.883	8.072
0.00					
	0.050	1.000	1.999	4.989	9.967
	0.200	1.000	1.992	4.938	9.850
	0.500	1.000	1.974	4.891	9.779
	0.800	1.000	1.952	4.833	9.671
0.10					
	0.050	1.000	2.057	5.241	10.541
	0.200	1.000	2.194	5.641	11.316
	0.500	1.000	2.416	6.173	12.295
	0.800	1.000	2.714	7.006	13.707
0.20					
	0.050	1.000	2.116	5.508	11.157
	0.200	1.000	2.418	6.474	13.103
	0.500	1.000	2.956	7.935	15.917
	0.800	1.000	3.719	10.334	20.449
0.50					
	0.050	1.000	2.305	6.410	13.283
	0.200	1.000	3.244	9.949	20.857
	0.500	1.000	5.418	17.218	35.943
	0.800	1.000	9.192	30.683	63.699
1.00					
	0.050	1.000	2.661	8.320	17.996
	0.200	1.000	5.402	21.666	49.080
	0.500	1.000	16.368	73.497	168.166
	0.800	1.000	58.477	280.705	643.200

Table 4.15

Inflation factor for estimator of odds ratio

Clusters of size 10

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.10					
	0.050	1.000	0.972	0.950	0.943
	0.200	1.000	0.905	0.873	0.869
	0.500	1.000	0.810	0.815	0.825
	0.800	1.000	0.703	0.777	0.807
0.00					
	0.050	1.000	1.000	0.998	0.997
	0.200	1.000	0.996	0.988	0.985
	0.500	1.000	0.987	0.978	0.978
	0.800	1.000	0.976	0.967	0.967
0.10					
	0.050	1.000	1.028	1.048	1.054
	0.200	1.000	1.097	1.128	1.132
	0.500	1.000	1.208	1.235	1.229
	0.800	1.000	1.357	1.401	1.371
0.20					
	0.050	1.000	1.058	1.102	1.116
	0.200	1.000	1.209	1.295	1.310
	0.500	1.000	1.478	1.587	1.592
	0.800	1.000	1.860	2.067	2.045
0.50					
	0.050	1.000	1.152	1.282	1.328
	0.200	1.000	1.622	1.990	2.086
	0.500	1.000	2.709	3.444	3.594
	0.800	1.000	4.596	6.137	6.370
1.00					
	0.050	1.000	1.331	1.664	1.800
	0.200	1.000	2.701	4.333	4.908
	0.500	1.000	8.184	14.699	16.817
	0.800	1.000	29.239	56.141	64.320

Table 4.16

Limiting value of the usual estimator

Clusters of size 25

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.04					
	0.050	0.000	0.666	1.569	2.260
	0.200	0.000	0.626	1.540	2.233
	0.500	0.000	0.605	1.530	2.225
	0.800	0.000	0.595	1.528	2.222
0.00					
	0.050	0.000	0.692	1.607	2.299
	0.200	0.000	0.690	1.605	2.298
	0.500	0.000	0.690	1.607	2.301
	0.800	0.000	0.692	1.609	2.302
0.10					
	0.050	0.000	0.759	1.703	2.402
	0.200	0.000	0.859	1.792	2.486
	0.500	0.000	0.962	1.866	2.554
	0.800	0.000	1.110	1.949	2.620
0.20					
	0.050	0.000	0.826	1.803	2.510
	0.200	0.000	1.035	2.002	2.703
	0.500	0.000	1.272	2.208	2.898
	0.800	0.000	1.603	2.502	3.167
0.50					
	0.050	0.000	1.029	2.113	2.849
	0.200	0.000	1.575	2.684	3.418
	0.500	0.000	2.198	3.307	4.038
	0.800	0.000	2.871	4.004	4.729
1.00					
	0.050	0.000	1.375	2.663	3.463
	0.200	0.000	2.535	3.940	4.755
	0.500	0.000	3.843	5.263	6.079
	0.800	0.000	5.209	6.630	7.446

Table 4.17

Asymptotic bias of the usual estimator

Clusters of size 25

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.04					
	0.050	0.000	-0.027	-0.040	-0.043
	0.200	0.000	-0.067	-0.069	-0.069
	0.500	0.000	-0.088	-0.079	-0.078
	0.800	0.000	-0.098	-0.082	-0.080
0.00					
	0.050	0.000	-0.001	-0.003	-0.003
	0.200	0.000	-0.003	-0.004	-0.005
	0.500	0.000	-0.003	-0.002	-0.002
	0.800	0.000	-0.002	-0.001	-0.001
0.10					
	0.050	0.000	0.066	0.094	0.100
	0.200	0.000	0.166	0.182	0.183
	0.500	0.000	0.269	0.257	0.252
	0.800	0.000	0.417	0.340	0.318
0.20					
	0.050	0.000	0.133	0.194	0.207
	0.200	0.000	0.342	0.393	0.400
	0.500	0.000	0.579	0.599	0.596
	0.800	0.000	0.910	0.892	0.864
0.50					
	0.050	0.000	0.336	0.503	0.547
	0.200	0.000	0.882	1.074	1.116
	0.500	0.000	1.505	1.698	1.735
	0.800	0.000	2.178	2.394	2.427
1.00					
	0.050	0.000	0.682	1.054	1.160
	0.200	0.000	1.841	2.331	2.453
	0.500	0.000	3.150	3.653	3.797
	0.800	0.000	4.516	5.020	5.144

Table 4.18

Relative bias of the usual estimator

Clusters of size 25

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.04					
	0.050		-0.039	-0.025	-0.019
	0.200		-0.097	-0.043	-0.030
	0.500		-0.126	-0.049	-0.034
	0.800		-0.141	-0.051	-0.035
0.00					
	0.050		-0.001	-0.002	-0.002
	0.200		-0.005	-0.003	-0.002
	0.500		-0.004	-0.001	-0.001
	0.800		-0.002	0.000	0.000
0.10					
	0.050		0.095	0.058	0.043
	0.200		0.240	0.113	0.080
	0.500		0.388	0.160	0.109
	0.800		0.601	0.211	0.138
0.20					
	0.050		0.192	0.120	0.090
	0.200		0.493	0.244	0.174
	0.500		0.835	0.372	0.259
	0.800		1.313	0.554	0.375
0.50					
	0.050		0.485	0.313	0.237
	0.200		1.273	0.668	0.485
	0.500		2.171	1.055	0.754
	0.800		3.142	1.488	1.054
1.00					
	0.050		0.984	0.655	0.504
	0.200		2.657	1.448	1.065
	0.500		4.544	2.270	1.640
	0.800		6.515	3.119	2.234

Table 4.19

Limiting value of estimator of odds ratio

Clusters of size 25

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.04					
	0.050	1.000	1.946	4.803	9.580
	0.200	1.000	1.871	4.665	9.332
	0.500	1.000	1.832	4.620	9.250
	0.800	1.000	1.814	4.608	9.227
0.00					
	0.050	1.000	1.999	4.987	9.965
	0.200	1.000	1.994	4.978	9.954
	0.500	1.000	1.994	4.989	9.979
	0.800	1.000	1.997	4.997	9.994
0.10					
	0.050	1.000	2.137	5.493	11.047
	0.200	1.000	2.361	6.000	12.014
	0.500	1.000	2.618	6.464	12.864
	0.800	1.000	3.034	7.024	13.740
0.20					
	0.050	1.000	2.285	6.068	12.302
	0.200	1.000	2.815	7.405	14.917
	0.500	1.000	3.568	9.099	18.146
	0.800	1.000	4.970	12.202	23.734
0.50					
	0.050	1.000	2.798	8.271	17.278
	0.200	1.000	4.832	14.642	30.520
	0.500	1.000	9.006	27.304	56.715
	0.800	1.000	17.652	54.813	113.210
1.00					
	0.050	1.000	3.957	14.343	31.911
	0.200	1.000	12.610	51.432	116.214
	0.500	1.000	46.668	192.984	436.695
	0.800	1.000	182.913	757.316	1713.712

157

Table 4.20

Inflation factor for estimator of odds ratio

Clusters of size 25

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.04					
	0.050	1.000	0.973	0.961	0.958
	0.200	1.000	0.935	0.933	0.933
	0.500	1.000	0.916	0.924	0.925
	0.800	1.000	0.907	0.922	0.923
0.00					
	0.050	1.000	0.999	0.997	0.997
	0.200	1.000	0.997	0.996	0.995
	0.500	1.000	0.997	0.998	0.998
	0.800	1.000	0.998	0.999	0.999
0.10					
	0.050	1.000	1.068	1.099	1.105
	0.200	1.000	1.181	1.200	1.201
	0.500	1.000	1.309	1.293	1.286
	0.800	1.000	1.517	1.405	1.374
0.20					
	0.050	1.000	1.142	1.214	1.230
	0.200	1.000	1.408	1.481	1.492
	0.500	1.000	1.784	1.820	1.815
	0.800	1.000	2.485	2.440	2.373
0.50					
	0.050	1.000	1.399	1.654	1.728
	0.200	1.000	2.416	2.928	3.052
	0.500	1.000	4.503	5.461	5.672
	0.800	1.000	8.826	10.963	11.321
1.00					
	0.050	1.000	1.978	2.869	3.191
	0.200	1.000	6.305	10.286	11.621
	0.500	1.000	23.334	38.597	43.670
	0.800	1.000	91.457	151.463	171.371

Table 4.21

Limiting value of the usual estimator

Clusters of size 50

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.02					
	0.050	0.000	0.671	1.582	2.274
	0.200	0.000	0.656	1.572	2.265
	0.500	0.000	0.652	1.570	2.263
	0.800	0.000	0.651	1.569	2.263
0.00					
	0.050	0.000	0.692	1.607	2.300
	0.200	0.000	0.692	1.608	2.301
	0.500	0.000	0.693	1.609	2.302
	0.800	0.000	0.693	1.609	2.302
0.10					
	0.050	0.000	0.803	1.744	2.441
	0.200	0.000	0.901	1.823	2.516
	0.500	0.000	0.965	1.867	2.557
	0.800	0.000	1.032	1.896	2.581
0.20					
	0.050	0.000	0.915	1.890	2.595
	0.200	0.000	1.137	2.087	2.786
	0.500	0.000	1.345	2.260	2.949
	0.800	0.000	1.649	2.491	3.155
0.50					
	0.050	0.000	1.256	2.357	3.093
	0.200	0.000	1.886	2.983	3.716
	0.500	0.000	2.545	3.646	4.377
	0.800	0.000	3.294	4.413	5.139
1.00					
	0.050	0.000	1.852	3.206	4.014
	0.200	0.000	3.218	4.617	5.430
	0.500	0.000	4.569	5.970	6.784
	0.800	0.000	5.944	7.346	8.159

Table 4.22

Asymptotic bias of the usual estimator

Clusters of size 50

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.02					
	0.050	0.000	-0.023	-0.028	-0.028
	0.200	0.000	-0.038	-0.037	-0.037
	0.500	0.000	-0.041	-0.040	-0.039
	0.800	0.000	-0.042	-0.040	-0.040
0.00					
	0.050	0.000	-0.001	-0.002	-0.002
	0.200	0.000	-0.001	-0.001	-0.001
	0.500	0.000	0.000	0.000	0.000
	0.800	0.000	0.000	0.000	0.000
0.10					
	0.050	0.000	0.109	0.135	0.139
	0.200	0.000	0.208	0.213	0.213
	0.500	0.000	0.272	0.257	0.255
	0.800	0.000	0.339	0.286	0.278
0.20					
	0.050	0.000	0.221	0.281	0.292
	0.200	0.000	0.444	0.478	0.483
	0.500	0.000	0.652	0.650	0.646
	0.800	0.000	0.956	0.881	0.853
0.50					
	0.050	0.000	0.563	0.747	0.790
	0.200	0.000	1.193	1.373	1.413
	0.500	0.000	1.852	2.036	2.075
	0.800	0.000	2.601	2.804	2.836
1.00					
	0.050	0.000	1.159	1.597	1.712
	0.200	0.000	2.525	3.007	3.127
	0.500	0.000	3.876	4.361	4.481
	0.800	0.000	5.251	5.736	5.856

Table 4.23

Relative bias of the usual estimator

Clusters of size 50

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.02					
	0.050		-0.033	-0.017	-0.012
	0.200		-0.054	-0.023	-0.016
	0.500		-0.059	-0.025	-0.017
	0.800		-0.060	-0.025	-0.017
0.00					
	0.050		-0.001	-0.001	-0.001
	0.200		-0.002	-0.001	-0.001
	0.500		-0.001	0.000	0.000
	0.800		0.000	0.000	0.000
0.10					
	0.050		0.158	0.084	0.060
	0.200		0.299	0.133	0.093
	0.500		0.392	0.160	0.111
	0.800		0.488	0.178	0.121
0.20					
	0.050		0.319	0.175	0.127
	0.200		0.641	0.297	0.210
	0.500		0.941	0.404	0.281
	0.800		1.380	0.548	0.370
0.50					
	0.050		0.813	0.464	0.343
	0.200		1.720	0.853	0.614
	0.500		2.672	1.265	0.901
	0.800		3.753	1.742	1.232
1.00					
	0.050		1.671	0.992	0.743
	0.200		3.643	1.868	1.358
	0.500		5.592	2.710	1.946
	0.800		7.576	3.564	2.543

Table 4.24

Limiting value of estimator of odds ratio

Clusters of size 50

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.02					
	0.050	1.000	1.955	4.864	9.720
	0.200	1.000	1.926	4.817	9.635
	0.500	1.000	1.919	4.806	9.613
	0.800	1.000	1.918	4.803	9.608
0.00					
	0.050	1.000	1.998	4.990	9.978
	0.200	1.000	1.998	4.993	9.986
	0.500	1.000	1.999	4.998	9.996
	0.800	1.000	2.000	4.999	9.999
0.10					
	0.050	1.000	2.231	5.722	11.488
	0.200	1.000	2.461	6.188	12.380
	0.500	1.000	2.625	6.468	12.900
	0.800	1.000	2.806	6.657	13.207
0.20					
	0.050	1.000	2.496	6.622	13.392
	0.200	1.000	3.119	8.065	16.208
	0.500	1.000	3.840	9.579	19.080
	0.800	1.000	5.204	12.072	23.464
0.50					
	0.050	1.000	3.513	10.557	22.034
	0.200	1.000	6.591	19.743	41.093
	0.500	1.000	12.744	38.313	79.606
	0.800	1.000	26.960	82.542	170.540
1.00					
	0.050	1.000	6.371	24.686	55.386
	0.200	1.000	24.990	101.156	228.119
	0.500	1.000	96.457	391.660	883.528
	0.800	1.000	381.569	1549.372	3494.705

Table 4.25

Inflation factor for estimator of odds ratio

Clusters of size 50

Values of β (followed by values of exp(β))

ρ_x	$\rho_{y/x}$	0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
-0.02					
	0.050	1.000	0.978	0.973	0.972
	0.200	1.000	0.963	0.963	0.964
	0.500	1.000	0.960	0.961	0.961
	0.800	1.000	0.959	0.961	0.961
0.00					
	0.050	1.000	0.999	0.998	0.998
	0.200	1.000	0.999	0.999	0.999
	0.500	1.000	1.000	1.000	1.000
	0.800	1.000	1.000	1.000	1.000
0.10					
	0.050	1.000	1.116	1.144	1.149
	0.200	1.000	1.231	1.238	1.238
	0.500	1.000	1.312	1.294	1.290
	0.800	1.000	1.403	1.331	1.321
0.20					
	0.050	1.000	1.248	1.324	1.339
	0.200	1.000	1.559	1.613	1.621
	0.500	1.000	1.920	1.916	1.908
	0.800	1.000	2.602	2.414	2.346
0.50					
	0.050	1.000	1.756	2.111	2.203
	0.200	1.000	3.295	3.949	4.109
	0.500	1.000	6.372	7.663	7.961
	0.800	1.000	13.480	16.508	17.054
1.00					
	0.050	1.000	3.185	4.937	5.539
	0.200	1.000	12.495	20.231	22.812
	0.500	1.000	48.229	78.332	88.353
	0.800	1.000	190.785	309.874	349.471

4.2.2 Two correlated binary covariates

It is known from section 3.3.3 that the usual estimator, β, is unbiased for the two covariate model when

$$\beta_1 = \beta_2 = 0.$$

Moreover, it was shown in that section that when

$$\beta_1 = 0,$$

but

$$\beta_2 \neq 0,$$

then two sufficient conditions for the unbiasedness of β_1 are

1. X_1 is distributed independently of X_2
2. there is no intracluster correlation, that is,

$$\rho_{y|x} = 0$$

However, no more information was obtained about the consistency, or lack of it, for β_2. Hence the size and sign of the bias of β_1 and β_2 are of interest. In this section we investigate this bias when the covariates are two correlated binary random variables.

For this investigation we use the method of limiting values of the estimating equations as developed in Secton 3.3.3. In particular, we have the equations (3.3.3.3)

$$E_X(F) = E_X(P)$$

$$E_X(\mathbf{X}F) = E_X(\mathbf{X}P)$$

where F is a function of the parameters of the correct (in this case, the Rosner) model, and P is a function of the parameters of the incorrect (in this case, the usual logistic) model. Expectation is taken in both cases with respect to the correct distribution. In the two covariate problem, these equations become

$$E_X(F) = E_X(P)$$
$$E_X(X_1F) = E_X(X_1P) \qquad (4.2.2.1)$$
$$E_X(X_2F) = E_X(X_2P)$$

These three non-linear equations may be solved iteratively to obtain the limiting values α_0, β_{10} and β_{20} of the estimators $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$.

When one of the parameters to be estimated, namely β_1, is zero, then F is a function of X_2 only, so that the equations may be written

$$E_{X_2}(F) = E_X(P)$$
$$E_{X_1}[X_1 E_{X_2|X_1}(F)] = E_X(X_1P) \qquad (4.2.2.2)$$
$$E_{X_2}(X_2F) = E_X(X_2P).$$

In order to evaluate the bias of the estimators we require a simple distribution for the X's, for example, a correlated binary distribution. We use Bahadur's(1961) representation of the correlated binary random variable (see section 1.4.3) with all second and higher order correlations set to 0, so that, for a random variable

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

where X_1 represents observations on the first unit, and X_2 represents observations on the second unit. Thus we have, as in (1.4.3.2)

$$P(x) = P_1(x)[1 + \sum_{j<k}^{4} E(Z_jZ_k)\, z_j\, z_k]$$

where

$$Z_j = (X_j - \mu_j)/\ \sigma_j, \quad j=1,\ldots,4,$$

and $P_1(x)$ is the usual binomial distiributon function for each unit, that is,

$$P_1(x) = \prod_{j=1}^{2} \binom{2}{x_{j+}} p_j^{x_{j+}} (1-p_j)^{(2-x_{j+})} \qquad (4.2.2.3)$$

where

$$p_j = \Pr(X_j=1)$$

and

$$x_{j+} = x_{j1} + x_{j2}$$

For simplicity we assume that

$$p_1 = p_2 = 0.5$$

so that (4.2.2.3) becomes

$$P_1(x) = \binom{2}{x_{1+}}\binom{2}{x_{2+}} / 16$$

The correlation structure may be written

$$\text{corr}(Z) = \begin{bmatrix} 1 & \rho_w & \rho_b & 0 \\ \rho_w & 1 & 0 & \rho_b \\ \rho_b & 0 & 1 & \rho_w \\ 0 & \rho_b & \rho_w & 1 \end{bmatrix}$$

where ρ_w represents correlaion within units but between covariates and ρ_b represents correlation bewteen units but within covariates.

Note that the model introduced here differs from the correlated binomial of Haseman and Kupper(1978) in that they considered only a one-dimensional model, that is, with one covariate representing scores on the different units within a cluster. In the model described above we are dealing with a two-dimensional model, that is, with scores on different units and on different variables. We refer to this model as the reduced correlated binary. Table 4.26 gives the probability distribution of the reduced correlated binary for all 16 possible values of $\mathbf{x}$, and table 4.27 gives some useful marginal distributions for this distributon, where $\mathbf{x}$ is written as $(x_{11}, x_{12}, x_{21}, x_{22})$ and x_{ij} is the observation on unit i, covariate j.

Having dealt with the distribution of X, we now specify the conditional distribution of Y. Let the usual logistic function be represented by

$$g(a) = \exp(a + t)/[1 + \exp(a + t)]$$

and Rosner's correlated logistic represented by

$$h(t_1, t_2) = [\exp(\alpha_1 + t_1) + \exp(\alpha_2 + t_2)] \; / \; [1 + \exp(\alpha_1 + t_1) + \exp(\alpha_1 + t_2) + \exp(\alpha_2 + t_1 + t_2)]$$

so that equations in (4.2.2.1) may be written

$$\sum_{\mathbf{x}_1, \mathbf{x}_2} h(\boldsymbol{\beta}'\mathbf{x}_1, \boldsymbol{\beta}'\mathbf{x}_2)\; p(x_{11}, x_{12}, x_{21}, x_{22}) = \sum_{x_1, x_2} g(\boldsymbol{\gamma}'\mathbf{x})\; p(x_1, x_2, +, +)$$

$$\sum_{x_{12}, x_{21}, x_{22}} h(\beta_1 + \beta_2 x_{12}, \beta_1 x_{21} + \beta_2 x_{22})\; p(1, x_{12}, x_{21}, x_{22}) = \sum_{x_2} g(\gamma_1 + \gamma_2 x_2)\; p(1, x_2, +, +)$$

$$\sum_{x_{11}, x_{21}, x_{22}} h(\beta_1 x_{11} + \beta_2, \beta_1 x_{21} + \beta_2 x_{22})\; p(x_{11}, 1, x_{21}, x_{22}) = \sum_{x_1} g(\gamma_1 x_1 + \gamma_2)\; p(x_1, 1, +, +)$$

Table 4.26

Values of the reduced correlated binary distribution

$x_{11}, x_{12}, x_{21}, x_{22}$	$16p(x)$
0,0,0,0	$1 + 2\rho_w + 2\rho_b$
1,0,0,0	
0,1,0,0	1
0,0,1,0	
0,0,0,1	
1,1,0,0	$1 + 2\rho_w - 2\rho_b$
1,0,1,0	$1 - 2\rho_w + 2\rho_b$
1,0,0,1	$1 - 2\rho_w - 2\rho_b$
0,1,1,0	$1 - 2\rho_w - 2\rho_b$
0,1,0,1	$1 - 2\rho_w + 2\rho_b$
0,0,1,1	$1 + 2\rho_w - 2\rho_b$
1,0,0,0	
0,1,0,0	1
0,0,1,0	
0,0,0,1	
1,1,1,1	$1 + 2\rho_w + 2\rho_b$

Table 4.27

Some marginal distributons based on the reduced correlated binary

$x_{11}, x_{12}, x_{21}, x_{22}$	$p(x)$
0,0,+,0	$(1 + \rho_w + \rho_b)/8$
0,0,+,1	$(1 + \rho_w - \rho_b)/8$
1,0,+,0	$(1 - \rho_w + \rho_b)/8$
1,0,+,1	$(1 - \rho_w - \rho_b)/8$
0,1,+,0	$(1 - \rho_w - \rho_b)/8$
0,1,+,1	$(1 - \rho_w + \rho_b)/8$
1,1,+,0	$(1 + \rho_w - \rho_b)/8$
1,1,+,1	$(1 + \rho_w + \rho_b)/8$
0,0,+,+	$(1 + \rho_w)/4$
1,0,+,+	$(1 - \rho_w)/4$
0,1,+,+	$(1 - \rho_w)/4$
1,1,+,+	$(1 + \rho_w)/4$
+,0,+,0	$(1 + \rho_b)/4$
+,0,+,1	$(1 - \rho_b)/4$
+,1,+,0	$(1 - \rho_b)/4$
+,1,+,1	$(1 + \rho_b)/4$

where $\boldsymbol{\beta}$ is the parameter of the correlated logistic and $\boldsymbol{\gamma}$ is the limiting value of the usual estimator.

When the coefficient of the first covariate is zero, that is, when

$$\beta_1 = 0,$$

then these equations may be written

$$\sum_{x_{12},x_{22}} h(\beta_2 x_{12},\beta_2 x_{22})\ p(+,x_{12},+,x_{22}) = \sum_{x_1,x_2} g(\boldsymbol{\gamma}'\mathbf{x})\ p(x_1,x_2,+,+)$$

$$\sum_{x_{12},x_{22}} h(\beta_2 x_{12},\beta_2 x_{22})\ p(1,x_{12},+,x_{22}) = \sum_{x_2} g(\gamma_1+\gamma_2 x_2)\ p(1,x_2,+,+)$$

$$\sum_{x_{22}} h(\beta_2,\beta_2 x_{22})\ p(+,1,+,x_{22}) = \sum_{x_1} g(\gamma_1 x_1+\gamma_2)\ p(x_1,1,+,+)$$

Next, using the reduced correlated binary distribution for the covariates, we may write these equations

$$[h(0,0) + h(\beta_2,\beta_2)](1+\rho_b) + [h(0,\beta_2) + h(\beta_2,0)](1-\rho_b)$$
$$= [g(0) + g(\gamma_1+\gamma_2)](1+\rho_w) + [g(\gamma_1) + g(\gamma_2)](1-\rho_w)$$

$$[h(0,0)(1-\rho_w+\rho_b) + h(\beta_2,\beta_2)(1+\rho_w+\rho_b)$$
$$+ h(\beta_2,0)(1+\rho_w-\rho_b) + h(0,\beta_2)(1-\rho_w-\rho_b)]$$
$$= 2[g(\gamma_1)(1-\rho_w) + g(\gamma_1+\gamma_2)(1+\rho_w)]$$

$$[h(\beta_2,0)(1-\rho_w) + h(\beta_2,\beta_2)(1+\rho_b)]$$
$$= [g(\gamma_2)(1-\rho_w) + g(\gamma_1+\gamma_2)(1+\rho_w)]$$

Finally, for simplicity of presentation of the results, we assume that

$$\rho_w = \rho_b = \rho_x,$$

and in tables 4.28 and 4.29 present values of the inflation factor of the estimator of the odds ratio related to the respective regression parameter, where the inflation factor was defined in secton 4.2.1 as the multiplier of the true odds ratio which yields the limiting value of the estimated odds ratio; for example,

$$\exp(\gamma_1)/\exp(\beta_1)$$

so that a value of the inflation factor less than one means a negatively-biased estimator, and a value greater than one means a positively-biased estimator.

Examination of tables 4.28 and 4.29 yield the following conclusions

1. the usual estimator of β_1 is always negatively biased.

2. the estimator of β_2 (which takes on various values) has the same pattern of bias as the estimator of the coefficient of a single covariate (see, for example, table 4.5), although for corresponding values of ρ_x and $\rho_{y|x}$, the bias is somewhat smaller in the two covariate case that in the single covariate case. The reduced bias may, however, be due to the reduced correlation structure in the reduced correlated binary distribution; for

Table 4.28

Values of inflation factor for bias of usual estimator of odds ratio for first of two binary covariates

ρ_x	$\rho_{y/x}$	$e^{\beta_2}(\beta_2)$: 1(0)	2(0.6931)	5(1.6094)	10(2.3026)
-0.80					
	0.00	1.000	1.000	1.000	1.000
	0.05	1.000	0.943	0.890	0.866
	0.20	1.000	0.792	0.633	0.565
	0.50	1.000	0.556	0.300	0.191
	0.80	1.000	0.378	(*)	(*)
-0.50					
	0.05	1.000	0.989	0.978	0.973
	0.20	1.000	0.957	0.916	0.897
	0.50	1.000	0.895	0.800	0.751
	0.80	1.000	0.837	0.683	0.589
0.00 (independent X's)					
	0.05	1.000	1.000	1.000	1.000
	0.20	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	1.000
	0.80	1.000	1.000	1.000	1.000
0.05					
	0.05	1.000	<1.00	<1.00	<1.00
	0.20	1.000	<1.00	0.999	0.999
	0.50	1.000	0.999	0.998	0.998
	0.80	1.000	0.999	0.997	0.996
0.10					
	0.05	1.000	<1.00	0.999	0.999
	0.20	1.000	0.999	0.997	0.997
	0.50	1.000	0.997	0.993	0.991
	0.80	1.000	0.995	0.989	0.986
0.50					
	0.05	1.000	0.989	0.978	0.973
	0.20	1.000	0.956	0.913	0.894
	0.50	1.000	0.892	0.790	0.750
	0.80	1.000	0.830	0.675	0.621
0.80					
	0.05	1.000	0.942	0.886	0.862
	0.20	1.000	0.783	0.597	0.531
	0.50	1.000	0.519	0.207	0.138
	0.80	1.000	0.310	(*)	(*)

(*) unique maximum likelihood estimate not attainable

Table 4.29

Values of inflation factor for bias of usual estimator of odds ratio for second of two binary covariates

ρ_x	$\rho_{y/x}$	$e^{\beta_2}(\beta_2)$: 1(0)	2(0.6931)	5(1.6094)	10(2.3026)
-0.80					
	0.00	1.000	1.000	1.000	1.000
	0.05	1.000	0.929	0.864	0.836
	0.20	1.000	0.746	0.561	0.489
	0.50	1.000	0.478	0.222	0.141
	0.80	1.000	0.294	(*)	(*)
-0.50					
	0.05	1.000	0.978	0.956	0.947
	0.20	1.000	0.915	0.834	0.798
	0.50	1.000	0.798	0.620	0.544
	0.80	1.000	0.692	0.439	0.335
0.00 (independent X's)					
	0.05	1.000	<1.00	<1.00	0.999
	0.20	1.000	0.999	0.994	0.990
	0.50	1.000	0.995	0.966	0.941
	0.80	1.000	0.989	0.918	0.860
0.05					
	0.05	1.000	1.002	1.003	1.004
	0.20	1.000	1.006	1.008	1.006
	0.50	1.000	1.012	1.000	0.982
	0.80	1.000	1.016	0.971	0.923
0.10					
	0.05	1.000	1.003	1.006	1.008
	0.20	1.000	1.013	1.022	1.023
	0.50	1.000	1.030	1.035	1.025
	0.80	1.000	1.044	1.029	0.992
0.50					
	0.05	1.000	1.022	1.045	1.056
	0.20	1.000	1.093	1.192	1.238
	0.50	1.000	1.248	1.549	1.694
	0.80	1.000	1.430	2.033	2.341
0.80					
	0.05	1.000	1.077	1.162	1.203
	0.20	1.000	1.354	1.884	2.184
	0.50	1.000	2.231	6.566	10.843
	0.80	1.000	4.091	(*)	(*)

(*) unique maximum likelihood estimate not attainable example, the correlation between scores on different units and different covariates is zero.

4.3 Bias of dummy variables estimator

It has been shown in section 3.3.4 that, for a fixed cluster size, the dummy variables estimator is, in general, a biased estimator of the regression parameter β. A sufficient for unbiasedness is that

$$\beta = 0,$$

that is, that all covariates have no effect.

However, the work of Pike, Hill and Smith(1980) with stratified and conditional estimators in case-control studies have shown that, athough the stratified estimator is badly biased in small strata, it becomes less biased in large strata, and, is asymptotically unbiased, that is, the bias becomes negligible as the cluster size(n) becomes large.

Hence it is possible that, for the correlated logistic model, in particular, for Rosner's version of the model, results similar to those obtained by Pike, Hill and Smith can be obtained, and that the dummy varibles estimator, which has very large bias for small cluster sizes, may become less biased for large clusters.

For a single binary covariate, the dummy variables estimator resembles the stratified estimator of Pike, Hill and Smith, and, as shown in section 2.4, the conditional estimator is the same as their conditional estimator.

The maximum likelihood equations given in (3.3.4.4) ad (3.3.4.5) can be written for a single covariate as

$$\mathbf{x}'\mathbf{z} = \mathbf{x}'\mathbf{p}$$

$$V'\mathbf{z} = V'\mathbf{p}$$

where $\mathbf{x}$, $\mathbf{z}$ and p are concatenations of the vectors

$$\mathbf{x}_i, \ \mathbf{z}_i, \ \mathbf{p}_i, \ i=1,...k,$$

where each vector represents the values in a cluster of the covariate, the response and estimated probability of response under the dummy variables model, respectively. The matrix V consists of k vectors, $\mathbf{v}_i$, each of length N(=kn), such that v_i contains 1's in positions (i-1)k to ik, and 0 elsewhere.

As indicated in section 3.3.4 these equations may be rewritten as

$$\mathbf{x}'\mathbf{z} = \mathbf{x}'\mathbf{p}$$

$$\mathbf{1}'\mathbf{z}_i = \mathbf{1}'\mathbf{p}_i, \ i=1,...,k. \quad (4.3.1)$$

Since the values of both $\mathbf{x}$ and $\mathbf{z}$ may be only 0 or 1, the first equation in (4.3.1) becomes

$$n_{11} = \sum_{x_{ij}=1} p_{ij}$$

where n_{11} was defined in the previous section as the number of units with both x_{ij} and z_{ij} equal to 1. Now p_{ij} in this equation is evaluated only for units with x equal to 1, so that we may replace p_{ij} in the equation with p_{1i}, where

$$p_{1i} = \exp(\gamma_i+\beta) / [1 + \exp(\gamma_i+\beta)]$$

Moreover, using m_{1i} to denote the number of units in cluster i with x equal to 1 we may write the equation as

$$n_{11} = \sum_{i=1}^{k} m_{1i}p_{1i}. \qquad (4.3.2)$$

The second set of equations in (4.3.1) may be written

$$m_{i1} = m_{0i}p_{0i} + m_{1i}p_{1i}, \quad i=1,\ldots,k \qquad (4.3.3)$$

where m_{i1} is the number of units in cluster i with y value equal to 1 and

$$p_{0i} = \exp(\gamma_i) / [1 + \exp(\gamma_i)]$$

We are excluding from equations (4.3.2) and (4.3.3) those clusters in which all y´s are 0 or 1. The reasons for this are explained in Appendix B.1, but it can be seen that, for example, that for a cluster in which all y´s are 1, equation (4.3.3) becomes

$$n = m_{0i}p_{0i} + m_{1i}p_{1i}$$

and, since

$$m_{0i} = n - n_{0i}$$

this may be written

$$n = np_{0i} + m_{1i}(p_{1i}-p_{0i})$$

which implies that

$$p_{0i} = p_{1i} = 1$$

which implies that

$$\gamma_i \longrightarrow \infty,$$

and there is no information about β in this cluster.

Similarly, for calculation of the limiting values of the estimators we can exclude those clusters with all values of x the same. For example, if in cluster i, we have all x values equal to 0; that is,

$$m_{0i} = 0,$$

then equation (3.3.3) becomes

$$m_1 = np_0$$

(where the subscript i has been omitted), and an explicit solution for q is obtained, namely

$$\gamma_i = \log\,[m_1/(n-m_1)]$$

(assuming we have already removed all clusters with $n=m_1$). Hence there is no information about parameter β in this cluster, and all clusters with all x's equal to 0 may be ignored in discussing limiting values of estimators of β. Similarly, it may be shown that all clusters with all x's equal to 1 may be ignored.

Now for all other values of n_{i1}, we write equation (4.2.3) as

$$n_{i1} = m_{0i}\,\frac{\exp(\gamma_i)}{1+\exp(\gamma_i)} + m_{i1}\,\frac{\exp(\gamma_i+\beta)}{1+\exp(\gamma_i+\beta)}$$

and we may solve for γ_i in terms of β. If we write c for $\exp(\gamma_i)$ and b for $\exp(\beta)$ and drop the subscript i, this equation becomes

$$n_1(1+c)(1+cb) = m_0c(1+cb) + m_1cb(1+c)$$

that is,

$$n_1 + n_1(1+b)c + n_1bc^2 = m_0c + m_0bc^2 + m_1bc + m_1bc^2$$

3

MICROCOPY RESOLUTION TEST CHART
NBS 1010a
ANSI and ISO TEST CHART No. 2

1.0 2.8 2.5 3.2 2.2 3.6 2.0 4.0 1.1 1.8 1.25 1.4 1.6

Transposing all terms to one side of the equation, and gathering like powers of c, we get

$$bc^2(m_0+m_1-n_1) + c(m_1 b+m_0-n_1-n_1 b) - n_1 = 0$$

which becomes

$$(n-n_1)bc^2 + [m_0-n_1+(m_1-n_1)b]c - n_1 = 0 \qquad (4.3.3)$$

which is a quadratic in c. Solving for c we get

$$c = \frac{-[m_0-n_1 + (m_1-n_1)b] \pm \sqrt{d}}{-2(n-n_1)\,b} \qquad (4.3.4)$$

where

$$d = [m_0-n_1+(m_0-n_1)b]^2-4(n-n_1)b(-n_1)$$

which, after some manipulation, becomes

$$d = (m_0-n_1)^2 + (m_1-n_1)b^2 + 2b[m_0 m_1+n_1 n-n_1^2] \qquad (4.3.5)$$

Next we must decide which root of equation (4.3.3) to select in order to yield non-negative values of $c(=\exp(q))$. The expression (4.3.4) may be written

$$c = [-e \pm (e^2-4fg)]/(-2f)$$

$$= [e \pm (e^2-4fg)]/2f \qquad (4.3.6)$$

where

$$e = m_0-n_1 + (m_1-n_1)b$$

$$f = (n-n_1)b$$

and

$$g = -n_1.$$

Now f is always positive (recall that all clusters with $n=n_1$ have been discarded), and g is always positive (recall that all clusters with $n_1=0$ are not used), so that

177

$$e^2 - 4fg > e^2$$

and

$$(e^2 - 4fg) > |e|$$

Hence, if e is positive, use of the negative square root in (4.3.6) will make c negative; hence, if e is positive, we must use the positive root. If e is negative, use of the negative square root in (4.3.6) makes c negative; however, use of the positive root will make c positive. Thus the solution of (4.3.3) for any γ is

$$\gamma = \log \{(m_0 - n_1) + (m_1 - n_1)b + [(m_0 - n_1)^2 + (m_1 - n_1)^2 b^2 + 2b(m_0 m_1 - n_1 + n_1^2)]^{1/2}\} / 2(n - n_1)b \qquad (4.3.7)$$

where

$$b = \exp(\beta).$$

This equation indicates that, for fixed n_1 and known β each γ is defined by a unique combination of m_1 and n_1. Hence there are at most $(n-1)^2$ unique values of γ, that is, $(n-1)^2$ unique clusters. Thus, in examining the bias in $\hat{\beta}_d$ as

$$k \longrightarrow \infty,$$

we may reduce the system of equations for the dummy variable estimates from k to $(n-1)^2$.

The next step is to examine the behaviour of equation (4.3.2) as

$$k \longrightarrow \infty.$$

This equation may be rewritten

$$\sum_{i=1}^{k} n_{11i} = \sum_{i=1}^{k} m_{1i}\, p_{1i}$$

where n_{11i} is the number of (x,y) pairs with value (1,1) in cluster i. Consider the limiting form of the average of this equation, as first proposed in section 3.3.3. We have

$$\lim_{k \to \infty} \frac{1}{k} \Sigma n_{11i} - \frac{1}{k} \Sigma m_{1i} p_{1i}$$

$$= E(R_{11}) - E(R_1 P)$$

where

$$R_1 = R_{10} + R_{11}$$

and R_{10} and R_{11} were defined in the previous section. Hence the limiting value of $\hat{\beta}_d$ is given by the solution to the equation

$$E(R_{11}) - E(R_1 P) = 0$$

plus $(n-1)^2$ equations of the form (4.3.7). Now

$$E(R_{11}) = \sum_{r=1}^{n} r \Pr(R_{11}=r)$$

where $\Pr(R_{11}=r)$ is a function of β, $\rho_{y|x}$, n, ρ_x and μ_x, where μ_x and ρ_x are the parameters of the underlying beta distribution. Similarly

$$E(R_1 P) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} (i+j) \frac{\exp(\gamma_{ij}+\beta)}{1+\exp(\gamma_{ij}+\beta)} \Pr(R_{11}=i, R_{10}=j)$$

where $\Pr(R_{11}=i, R_{10}=j)$ is a function of β, $\rho_{y|x}$, n, ρ_x and μ_x, and the subscript ij should more properly be written (j-1)i + j.

The program BIASD (see Appendix D.2) was written to evaluate the bias of the dummy variales estimator for various values of β, $\rho_{y|x}$, n and ρ_x. In all cases it was assumed that the underlying beta distribution has mean 0.5. The distributions used for Z and X were those used in the previous section, that is, Rosner's correlated logistic for Z (truncated, of course, not to allow all dependent variables of the same value within a cluster) and the beta-binomial for x.

Table 4.30 gives the value of β_0, the limiting value of $\hat{\beta}_d$, the bias and relative bias of β_0, the limiting value of the odds ratio estimator based on the dummy variables estimator, $\exp(\beta_0)$, and the inflation factor for $\exp(\beta_0)$, for various values of β, $\rho_{y|x}$ and ρ_x and for cluster sizes

$$n = 2,3,4\ 10,25,50.$$

From the table it can be seen that

1. there is no effect of the correlation structure, namely, ρ_x and $\rho_{y|x}$, on the bias as the cluster size changes.

2. the bias and relative bias increase with the absolute value of β, although the relative bias increases only a small amount.

3. the bias and relative bias decrease as n increases, but only at a slow rate. For example, the relative bias is slightly under 5 percent for clusters of size 25, but is still greater than 2 percent for size 50.

Table 4.30

Summary of measures of bias of dummy variables estimator for single binary covariate

Cluster size	Values of β (followed by values of exp(β)) 0.000 1.0	0.693 2.0	1.609 5.0	2.303 10.0
	Limiting value of dummy variables estimator			
2	0.000	1.386	3.219	4.605
3	0.000	1.050	2.539	3.783
5	0.000	0.871	2.066	3.035
10	0.000	0.772	1.805	2.599
25	0.000	0.723	1.682	2.408
50	0.000	0.708	1.645	2.354
	Asymptotic bias of dummy variables estimator			
2	0.000	0.693	1.609	2.303
3	0.000	0.357	0.929	1.481
5	0.000	0.178	0.456	0.732
10	0.000	0.079	0.196	0.297
25	0.000	0.029	0.072	0.106
50	0.000	0.014	0.035	0.051
	Relative bias of dummy variables estimator			
2		1.000	1.000	1.000
3		0.515	0.577	0.643
5		0.256	0.283	0.318
10		0.113	0.121	0.129
25		0.042	0.045	0.046
50		0.021	0.022	0.022
	Limiting value of estimator of odds ratio			
2	1.000	4.000	25.000	100.000
3	1.000	2.858	12.665	43.958
5	1.000	2.389	7.890	20.793
10	1.000	2.164	6.080	13.457
25	1.000	2.060	5.374	11.117
50	1.000	2.029	5.179	10.524
	Inflation factor for estimator of odds ratio			
2	1.000	2.000	5.000	10.000
3	1.000	1.429	2.533	4.396
5	1.000	1.194	1.578	2.079
10	1.000	1.082	1.216	1.346
25	1.000	1.030	1.075	1.112
50	1.000	1.015	1.036	1.052

The decrease in bias of the dummy variables estimator mirrors the decrease detected by Pike, Hill and Smith in their study, but it is not so dramatic with the correlated logistic model as in their study of the stratified logistic model.

4.4 Asymptotic variance and relative efficiency

4.4.1 No effect of covariates

Table 3.1 at the end of chapter 3 gives formulae for the asymptotic variance of the four estimators when

$$\beta = 0$$

in the case of p covariates. Where there is only one covariate, the table reduces to Table 4.31 given below. The corresponding relative efficiencies are given in Table 4.32.

Table 4.31

Asymptotic variance of estimators (one covariate, $\beta=0$)

Estimator	Asymptotic variance
maximum likelihood	$\{n\,\sigma_y^2\,\sigma_x^2\,[1 + (n-1)\,\rho_y\,\rho_x]\}^{-1}$
conditional	$[(n-1)\,\sigma_y^2\,\sigma_x^2\,(1-\rho_y)\,(1-\rho_x)]^{-1}$
usual	$[1 + (n-1)\,\rho_y\,\rho_x]/(n\,\sigma_y^2\,\sigma_x^2)$
dummy variables	$\dfrac{n}{(n-1)^2}\;\dfrac{[1 + (n-1)\,\rho_z\,\rho_x]\,c}{\sigma_z^2\,\sigma_x^2\,(1-\rho_x)^2}$

Table 4.32

Relative efficiency of estimators (one covariate, $\beta=0$)

Estimator	Relative efficiency
conditional	$\dfrac{(n-1)(1-\rho_y)(1-\rho_x)}{n[1+(n-1)\rho_y\rho_x]}$
usual	$[1+(n-1)\rho_y\rho_x]^{-2}$
dummy variables	$\dfrac{(n-1)^2\sigma_z^2(1-\rho_x)^2}{n^2\sigma_y^2[1+(n-1)\rho_y\rho_x][1+(n-1)\rho_z\rho_x]}$

From table 4.32, it may be seen that

1. the conditional estimator is never more efficient than the maximum likelihood estimator, and only attains the same efficiency when

$$\rho_x = -1/(n-1)$$

Proof:

Consider the denominator minus the numerator of the relative efficiency given in the table

$$n[1+(n-1)\rho_y\rho_x] - (n-1)(1-\rho_y)(1-\rho_x)$$

$$= n + n(n-1)\rho_y\rho_x - (n-1) + (n-1)\rho_y + (n-1)\rho_x - (n-1)\rho_y\rho_x$$

$$= 1 + (n-1)^2\rho_y\rho_x + (n-1)\rho_y + (n-1)\rho_x$$

$$= [1+(n-1)\rho_y][1+(n-1)\rho_x]$$

Since ρ_y is always non-negative, $[1+(n-1)\rho_y]$ is never less than

1, and since ρ_x, like any intraclass coefficient, is bounded below by $-1/(n-1)$, then this expression is always positive and is only equal to 0 when ρ_x attains its lower bound.

2. the usual estimator is more efficient than the maximum likelihood estimator when ρ_x is negative and ρ_y is non-zero (ρ_y is always non-negative).

3. for fixed n, the relative efficiency of the usual estimator is bounded above by the line $(1-\rho_y)^{-2}$. The maximum efficiency is attained when

$$\rho_x = -1/(n-1).$$

4. for large cluster size and non-negative values of ρ_x, the dummy variables estimator is never more efficient than the maximum likelihood estimator and attains maximum efficiency when ρ_x and ρ_y reach their lower bound of 0.

Proof:

For large cluster size, the difference between the random variables Y and Z becomes small (see, for example, Appendix B.2), so that

$$\sigma_z^2 = \sigma_y^2$$

and

$$\rho_z = \rho_y.$$

Moreover c is approximately 1 so that the relative efficiency of the dummy variables estimator may be written as

$$\left[\frac{(n-1)\ (1-\rho_x)}{n[1 + (n-1)\rho_y \rho_x]} \right]^2$$

Consider only the fraction in this expression and subtract the numerator from the denominator to give

$$n + n(n-1)\rho_y \rho_x - n + 1 + (n-1)\rho_x$$
$$= 1 + (n-1)\rho_x\ (1 + n\rho_y)$$

Now ρ_y is always non-negative so that $(1+n\rho_y)$ is never less than 1, and ρ_x is assumed to be non-negative so that $(n-1)\rho_x$ is bounded below by 0. This expression is monotone in ρ_y and ρ_x so that its minimum value of 1 is attained when both ρ_y and ρ_x are 0.

Figures 4.1 to 4.20 give plots of the relative efficiencies of the estimators for various values of cluster size(n) and ρ_x, plotted as a function of ρ_y.

Figures 4.1 to 4.6 give the relative efficiency of the conditional estimator for cluster sizes

$$n = 2,3,5,10,25,50$$

and for various values of ρ_x which include

$$\rho_x = 0.5,0.1,0.0$$

and at least one negative value of ρ_x. From these figures it may be seen that

Figure 4.1. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of
(n=2, single binary covariate, = 0)

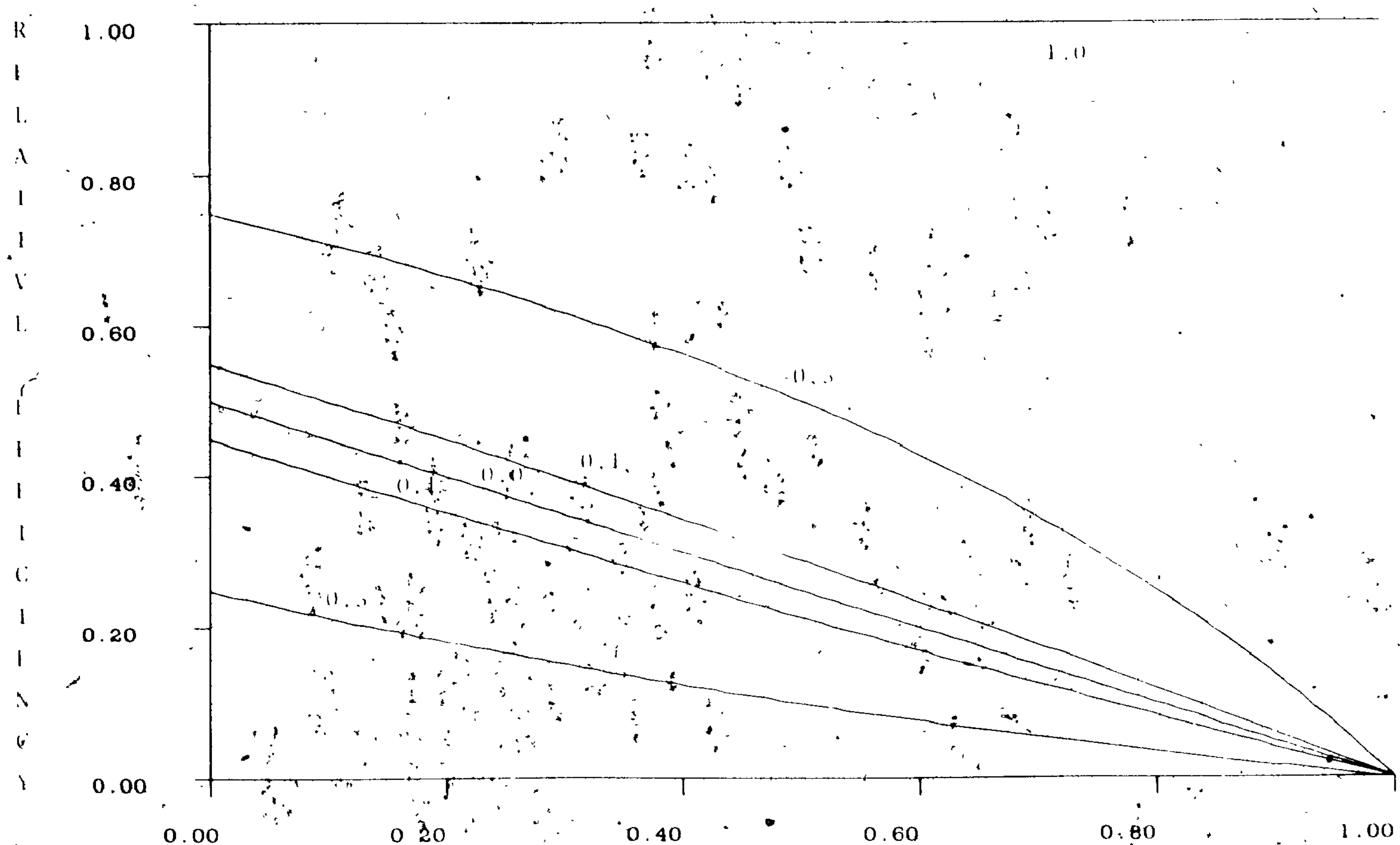


Figure 1.2. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of
(n 5, single binary covariate, 0)

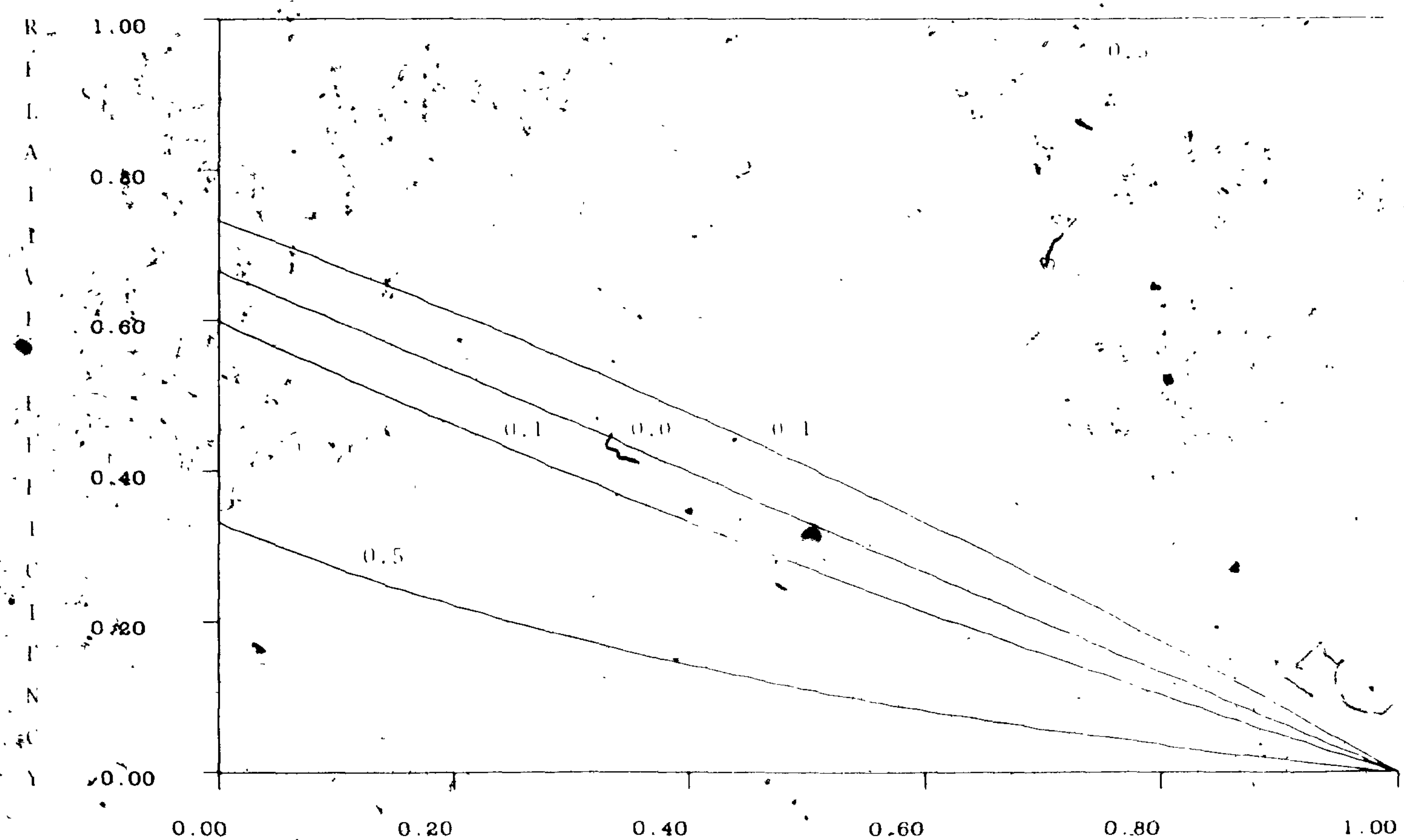


Figure 4.5. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of
(n 5, single binary covariate, : 0)

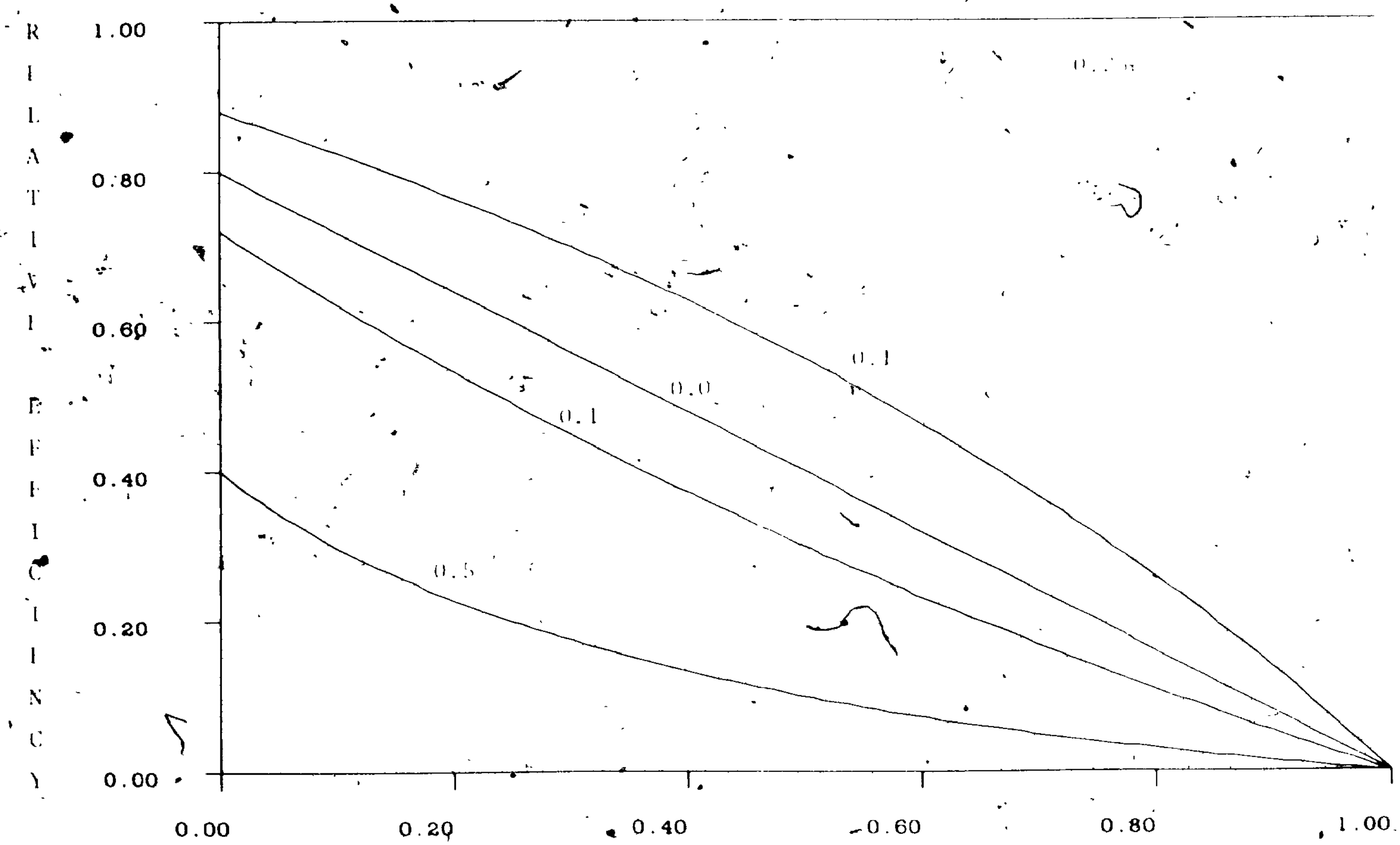


Figure 4.4. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of

(n=10, single binary covariate, 01)

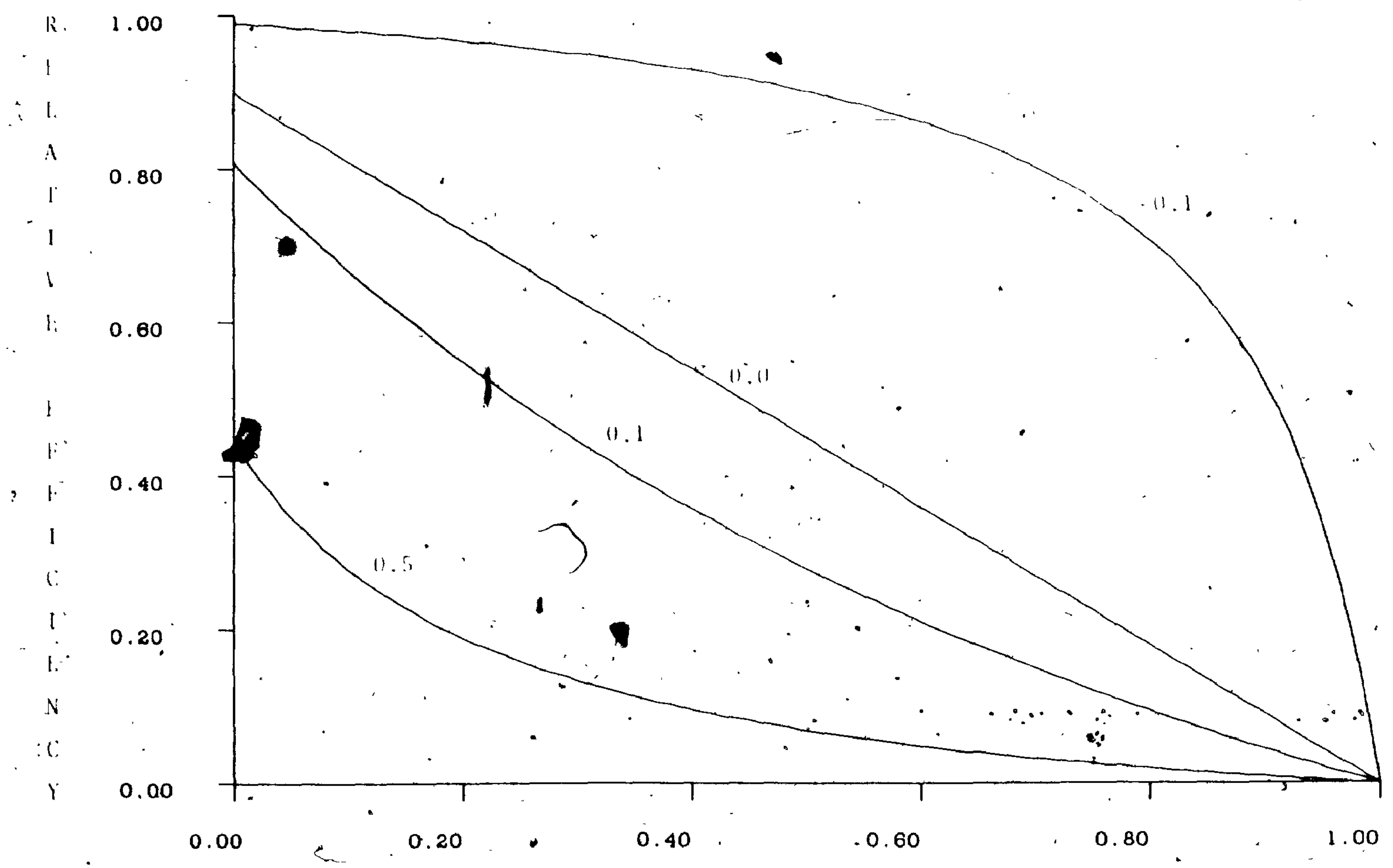


Figure 4.5. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of ρ_x (n=25, single binary covariate, ρ=0)

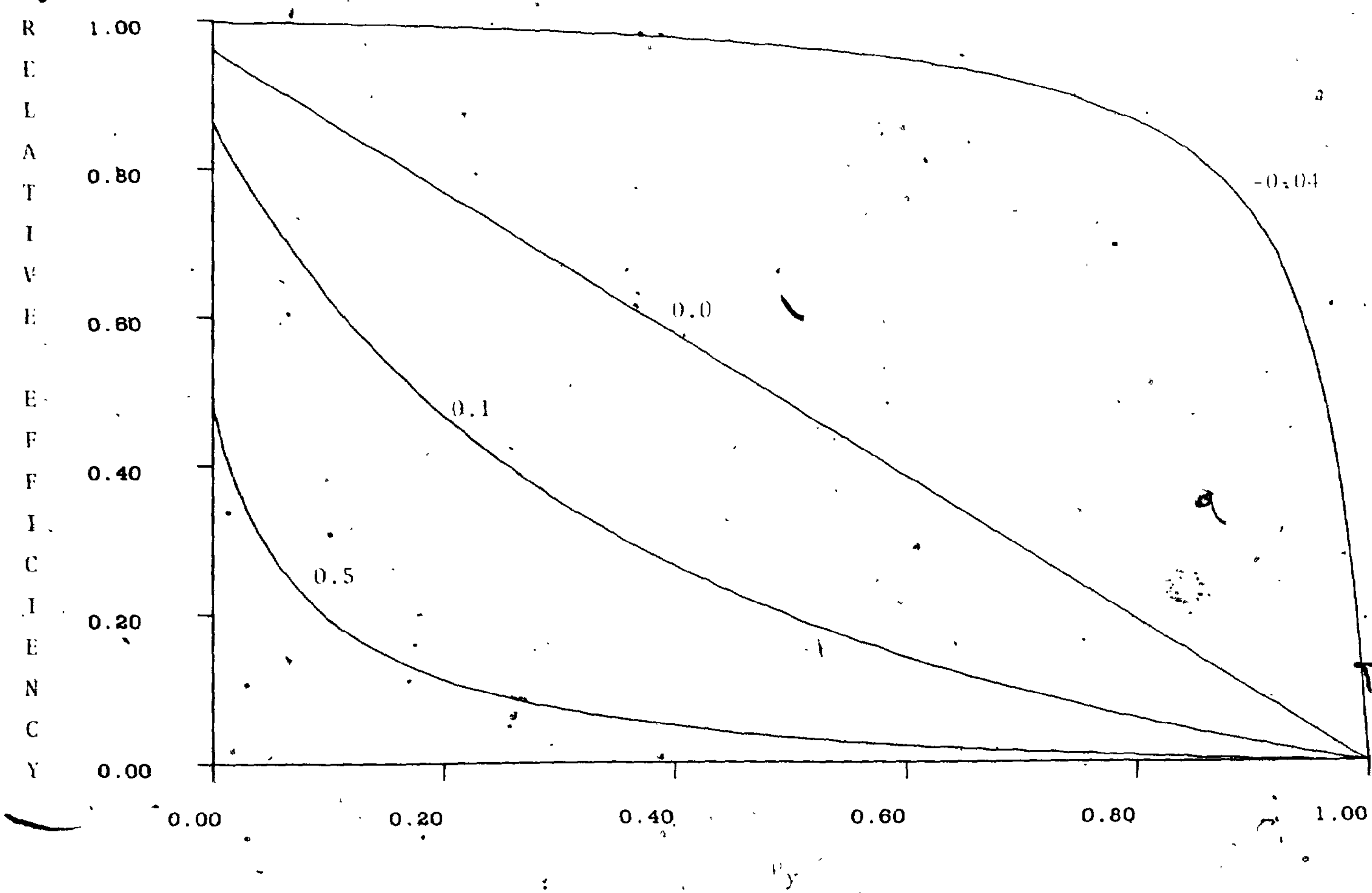
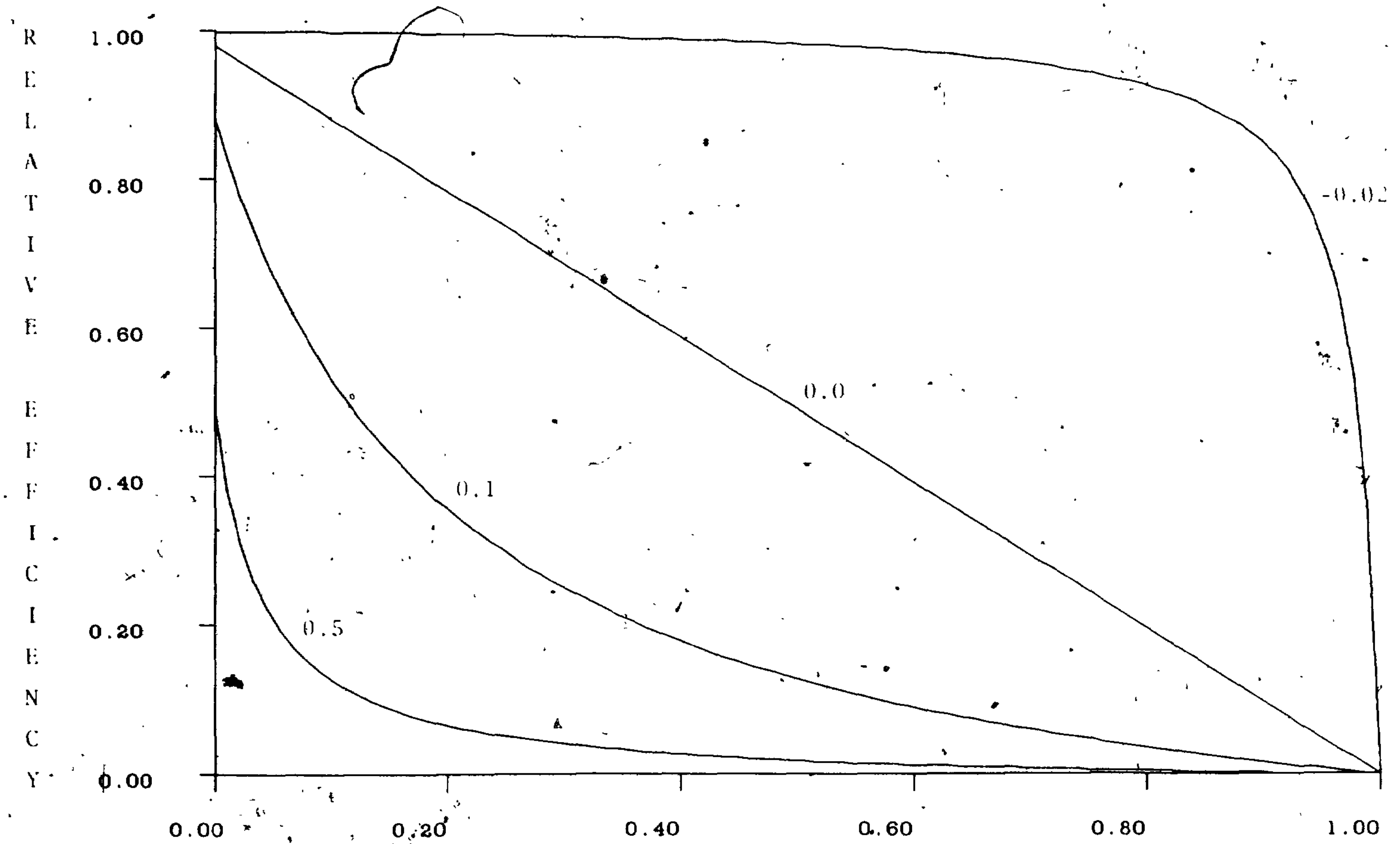


Figure 4.6. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of
(n=50, single binary covariate, r=0)



1. there is always one value of ρ_x, namely,

$$\rho_x = -1/(n-1)$$

for which the conditional is fully efficient,

2. the relative efficiency is a non-increasing function of ρ_y and ρ_x,

3. for non-positive values of ρ_x, for a fixed value of ρ_y, the relative efficiency increases with n.

4. for positive values of ρ_x, relative efficiency increases with n for values of ρ_y near 0, but decreases with n for values of ρ_y near 1.

Figures 4.7 to 4.10 give graphs of the relative efficiency of the usual estimator for

$$n = 2,5,10,25$$

and for non-negative values of ρ_x, namely,

$$\rho_x = 1.0, 0.5, 0.1, 0.0.$$

From these graphs it may be seen that, except for

$$\rho_x = 0,$$

the relative efficiency decreases with increasing values of ρ_y, ρ_x and n.

Figure 4.7. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-negative Values of
(n=2, single binary covariate, ß=0)

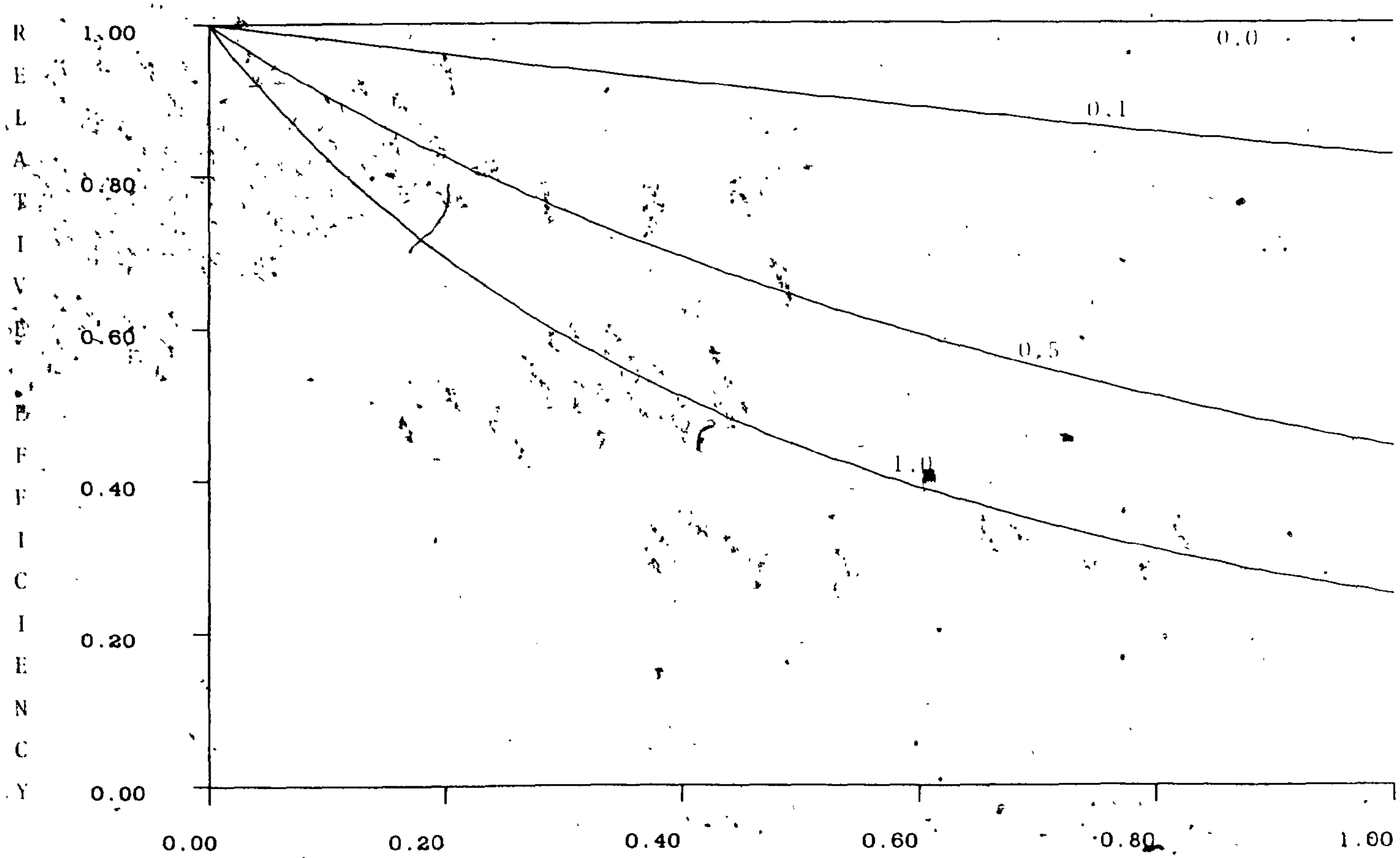


Figure 4.8. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-negative Values of
(n=5, single binary covariate, =0)

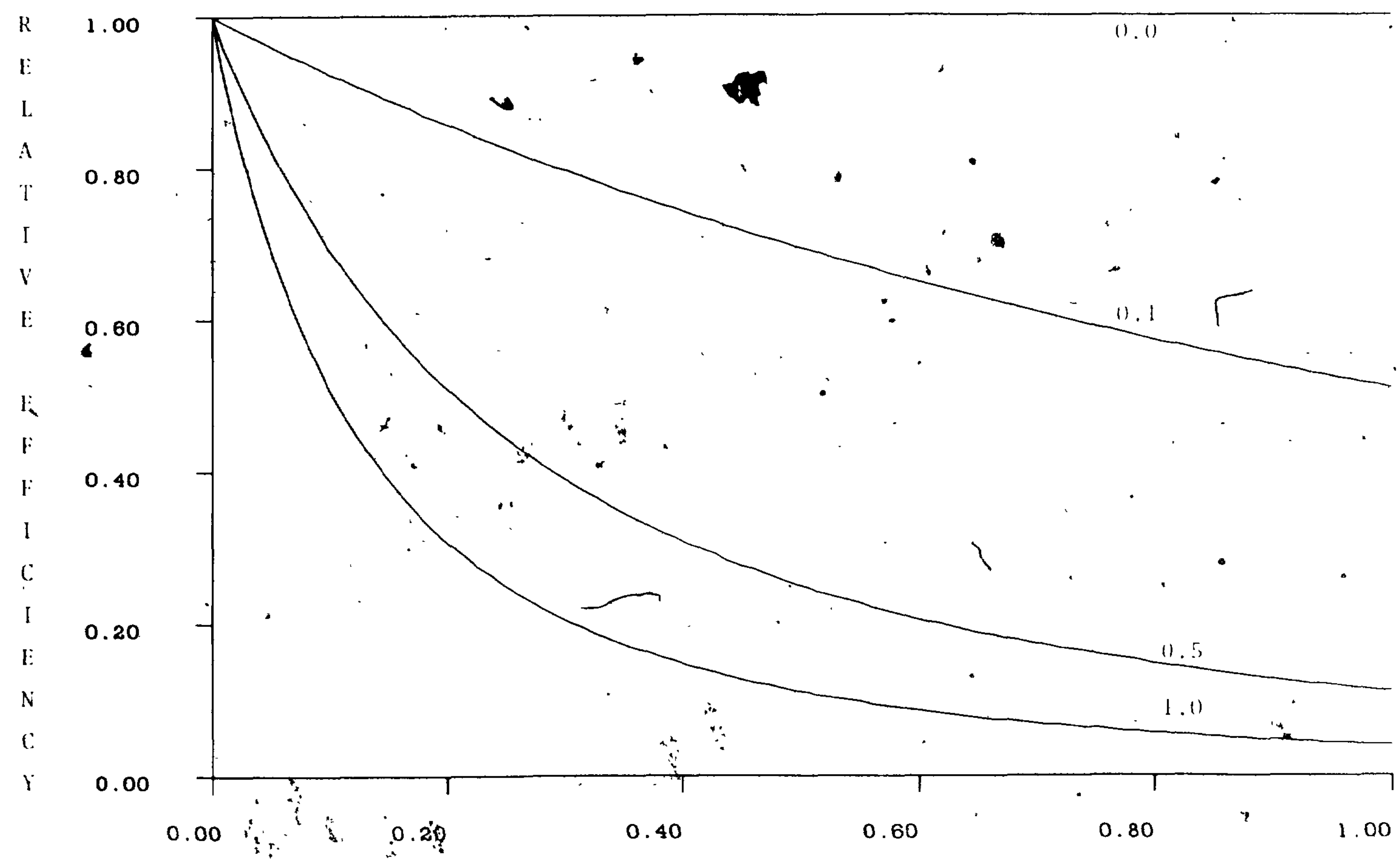


Figure 4.9. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-negative Values of
(n=10, single binary covariate, β=0)

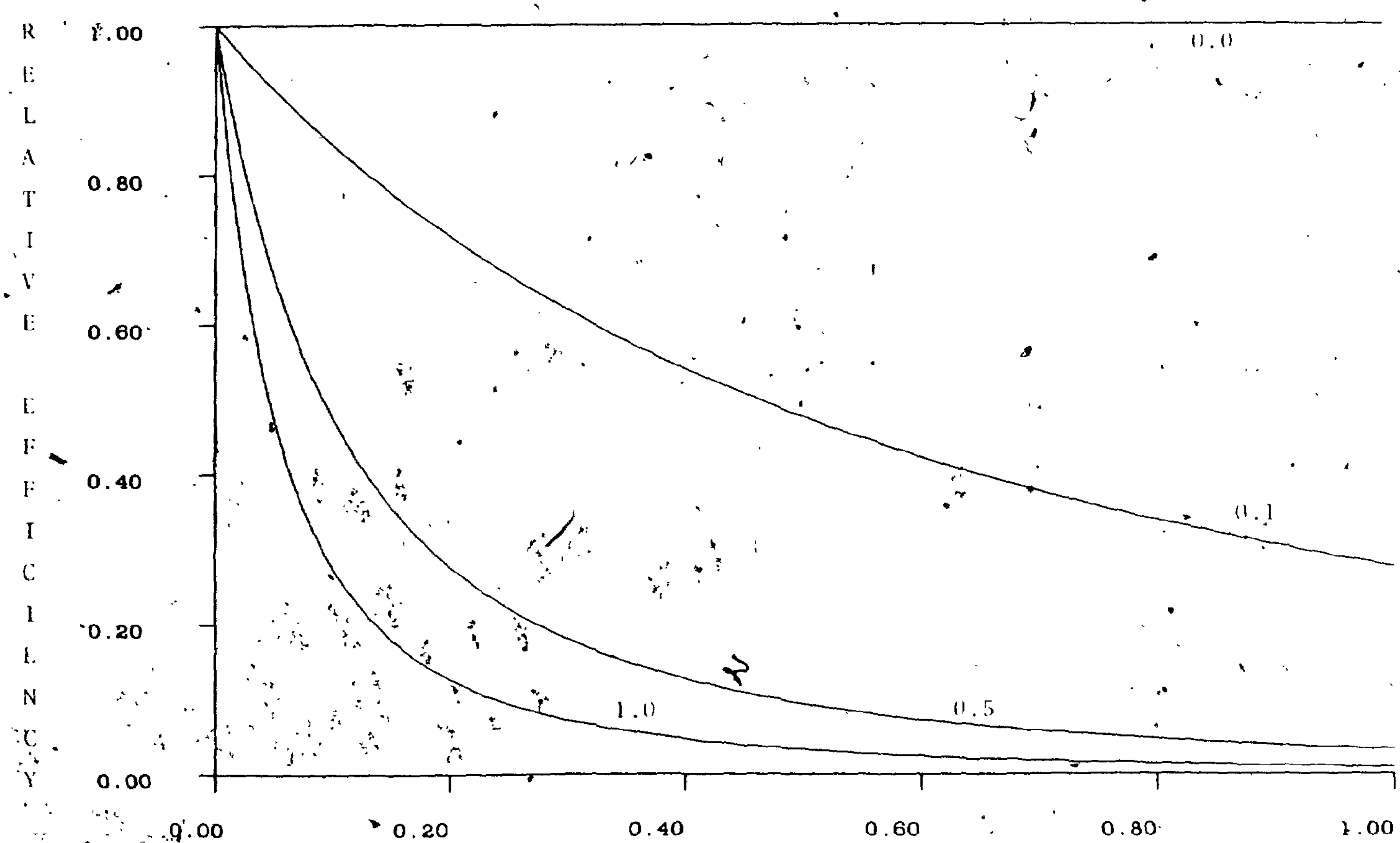
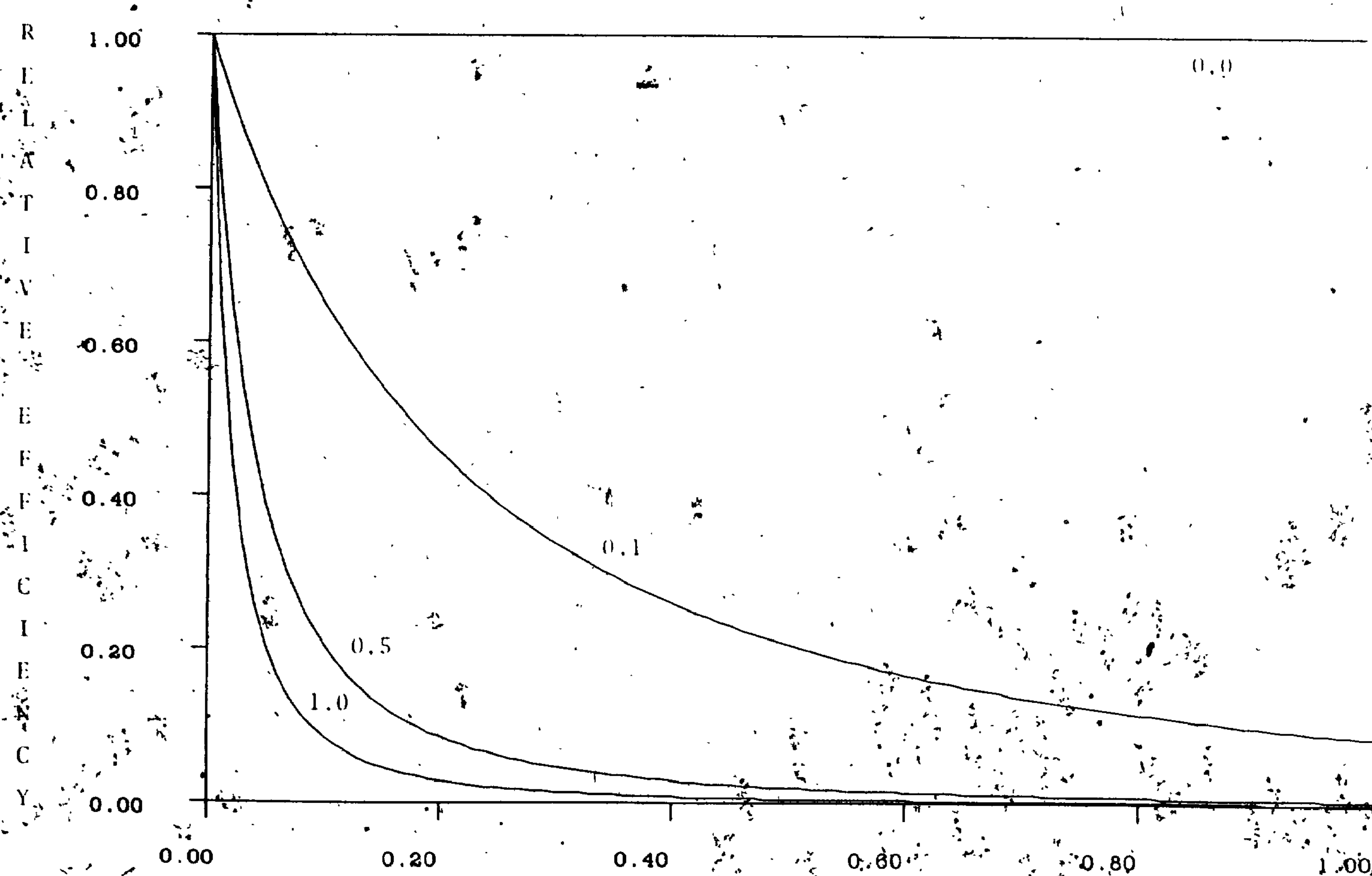


Figure 4.10. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-negative Values of λ (n=25, single binary covariate, r=0)



Figures 4.11 to 4.14 give graphs of the relative efficiency of the usual estimator for

$$n = 2,5,10,25$$

and for non-positive values of ρ_x, that is,

$$\rho_x = 0.0$$

plus at least two other negtive values of ρ_x. From these graphs it may be seen that, except for

$$\rho_x = 0$$

the relative efficiency increases with increasing values of ρ_y, ρ_x and n, with the limiting value of ρ_x being $-1/(n-1)$.

Figures 4.15 to 4.20 show the relative efficiency of the dummy variables estimator for

$$n = 2,3,5,10,25,50$$

and a range of values of ρ_x. From these figures it may be seen that

1. the relative efficiency at

$$n = 2$$
$$\beta = 0$$

   is exactly 1/4 that of the conditional estimator with the same parameter values(see Figure 4.1). The reason for this is explained below.

2. the relative efficiency increases, in general, with values of n, reaching 0.9 for

$$\rho_x = 0,$$

Figure 4.11. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-positive Non-parametric Values of λ

($n=2$, single binary covariate, $\gamma=0$)

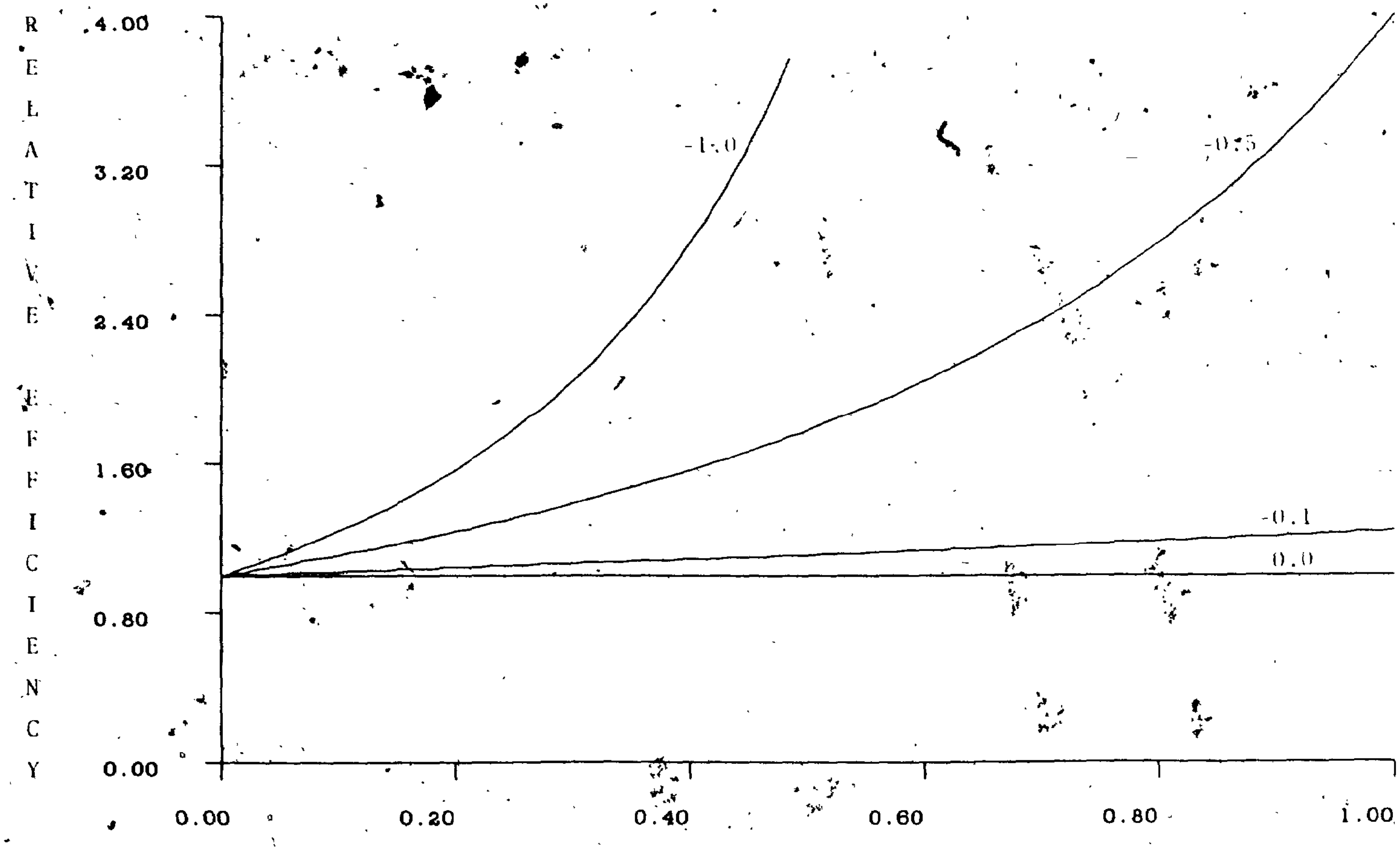


Figure 4.12. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-positive Non-parametric Values of
(n=5, single binary covariate, ρ=0)

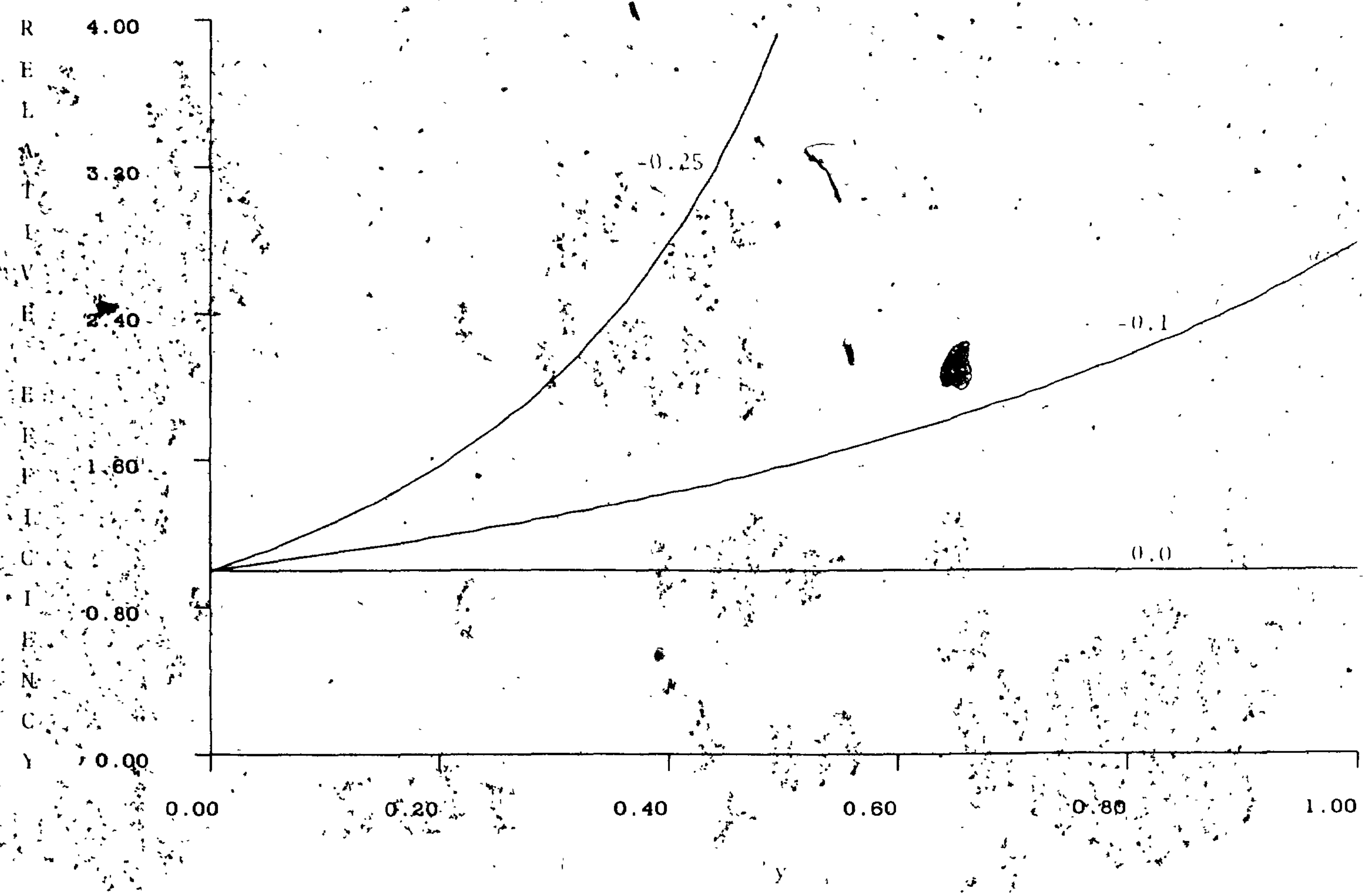


Figure 4.13. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-positive Non-parametric Values of

(n=10, single binary covariate, =0)

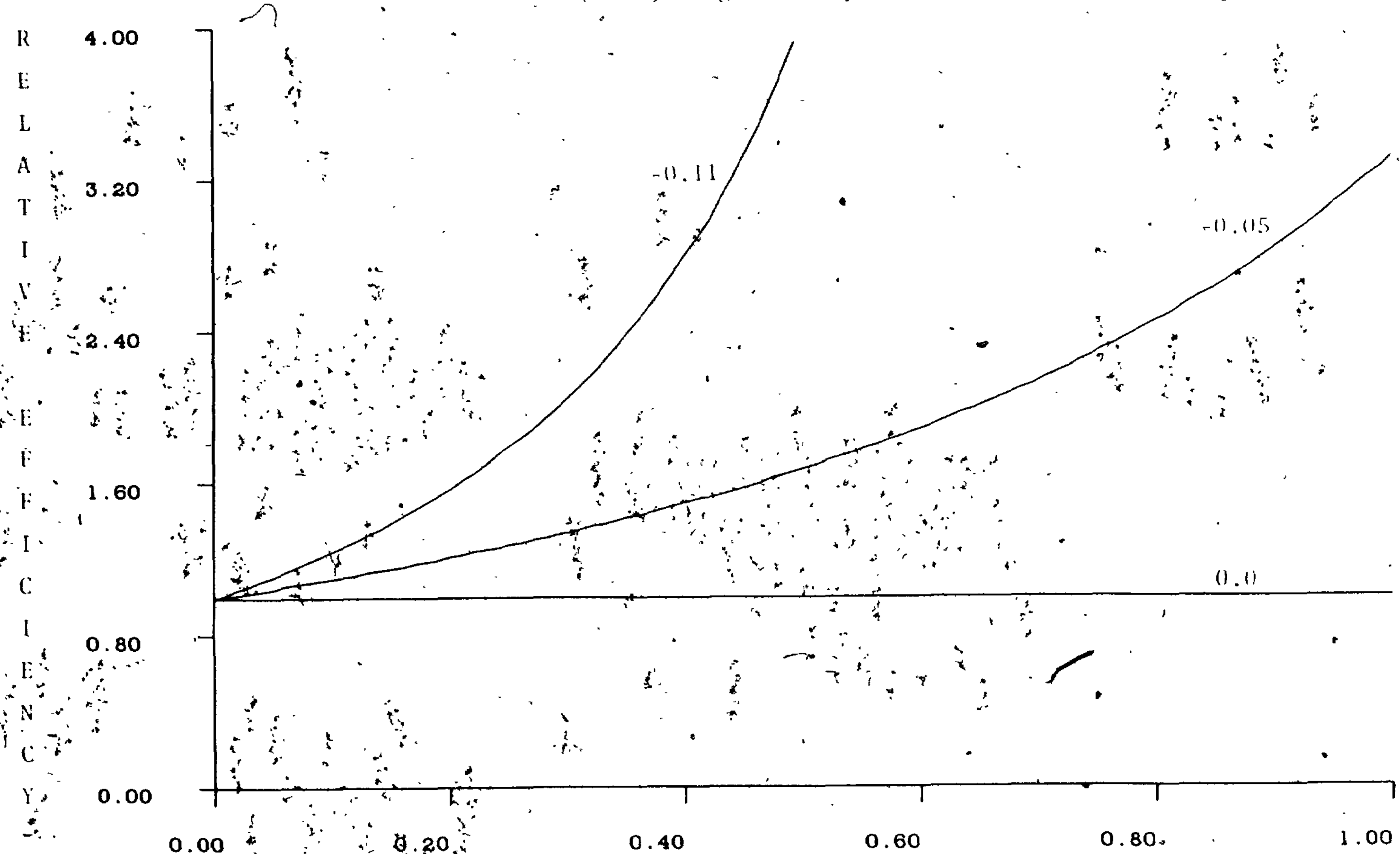
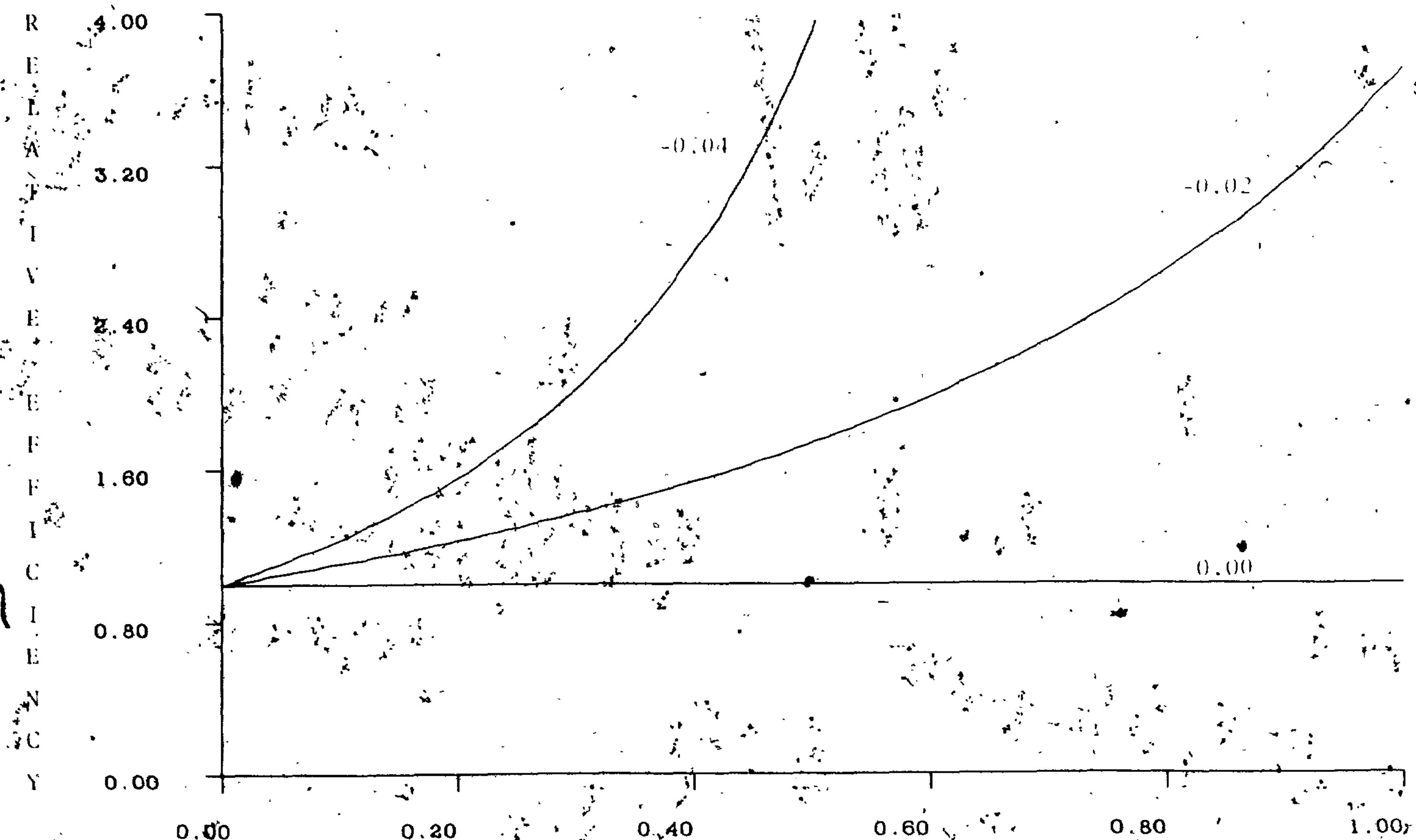


Figure 4.14. Asymptotic Relative Efficiency of the Usual Estimator for Varying Non-positive Non-parametric Values of λ

(n=25, single binary covariate, r=0)



and about 1.0 for negative values of ρ_x.

3. the relative efficiency declines with increasing values of ρ_y and ρ_x, except for negative values of ρ_x near the boundary of the parameter space.

Finally it should be noted that the relationship between the asymptotic variances of the conditional and the the dummy variables estimators given by Table 4.31 agrees, in the case of clusters of size 2, with another approach to the problem. When n is 2 and we assume that the underlying beta distribution has mean 0.5, then Y is a binomial random variable with parameters 1 and 0.5, so that

$$\sigma_y^2 = 0.25$$

Now Z is the random vector which can take on values of (1,0) or (0,1) with equal probabilities, so that the probability distribution of Z is

$$f(0,1) = 0.5$$
$$f(1,0) = 0.5$$

(see for example Appendix B.2) so that

$$\sigma_z^2 = 0.25,$$
$$\rho_z = \mathrm{corr}(Z_1, Z_2) = [0 - (0.5)^2] / (0.5)^2$$
$$= -1$$

and

$$c = 1/ [1 - \Pr(Y_1 = Y_2 = 0) - \Pr(Y_1 = Y_2 = 1)]$$
$$= 1 / [0.5(1-\rho_y)]$$

so that the asymptotic variance of the dummy variables estimator is

$$2\ (1-\rho_x) / [0.5(1-\rho_y)\ 0.25\ \sigma_x^2\ (1-\rho_x)^2]$$

Figure 4.15. Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of [illegible]
(n=2, single binary covariate, [illegible]=0)

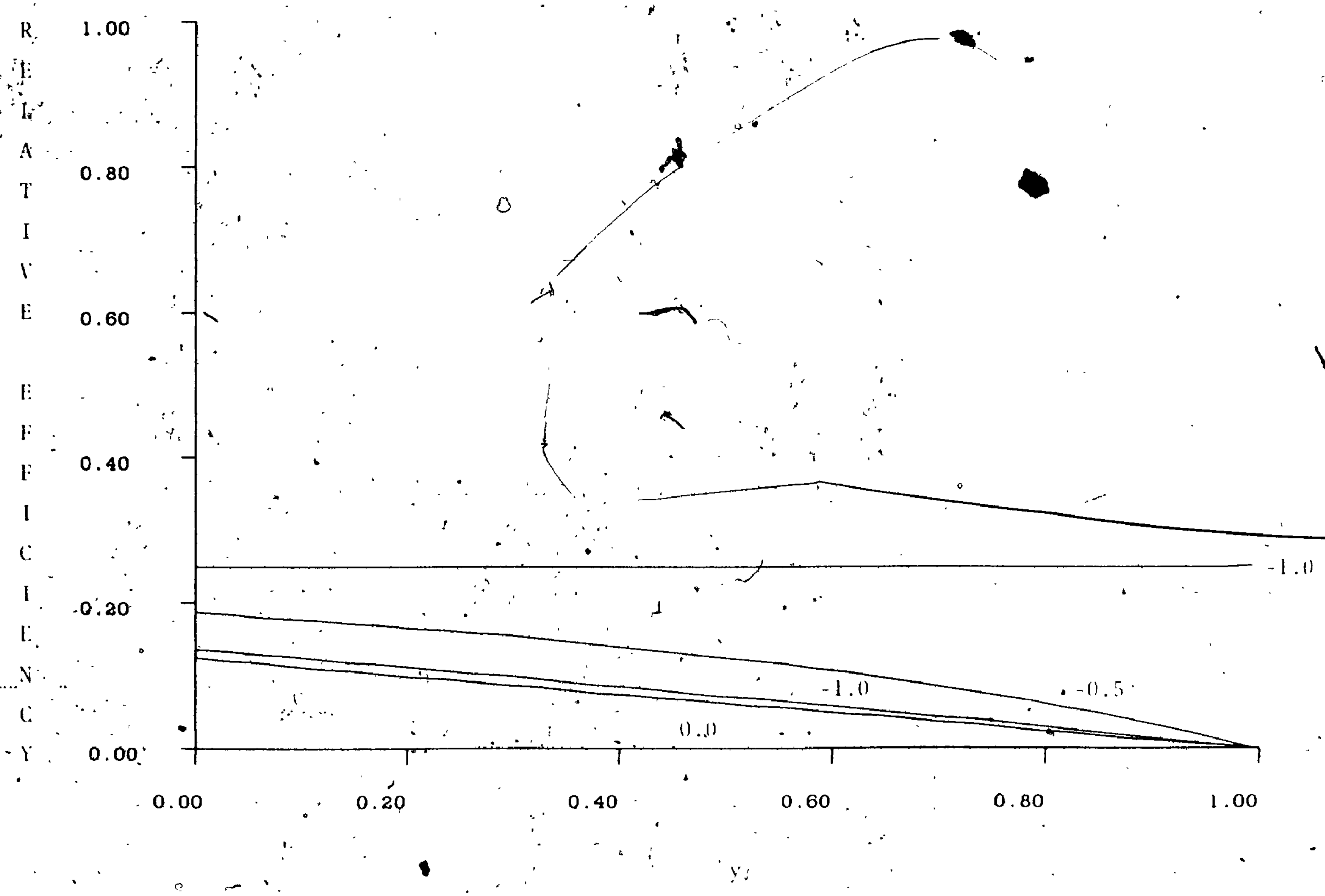


Figure 4.16. Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of
(n=3, single binary covariate, β=0)

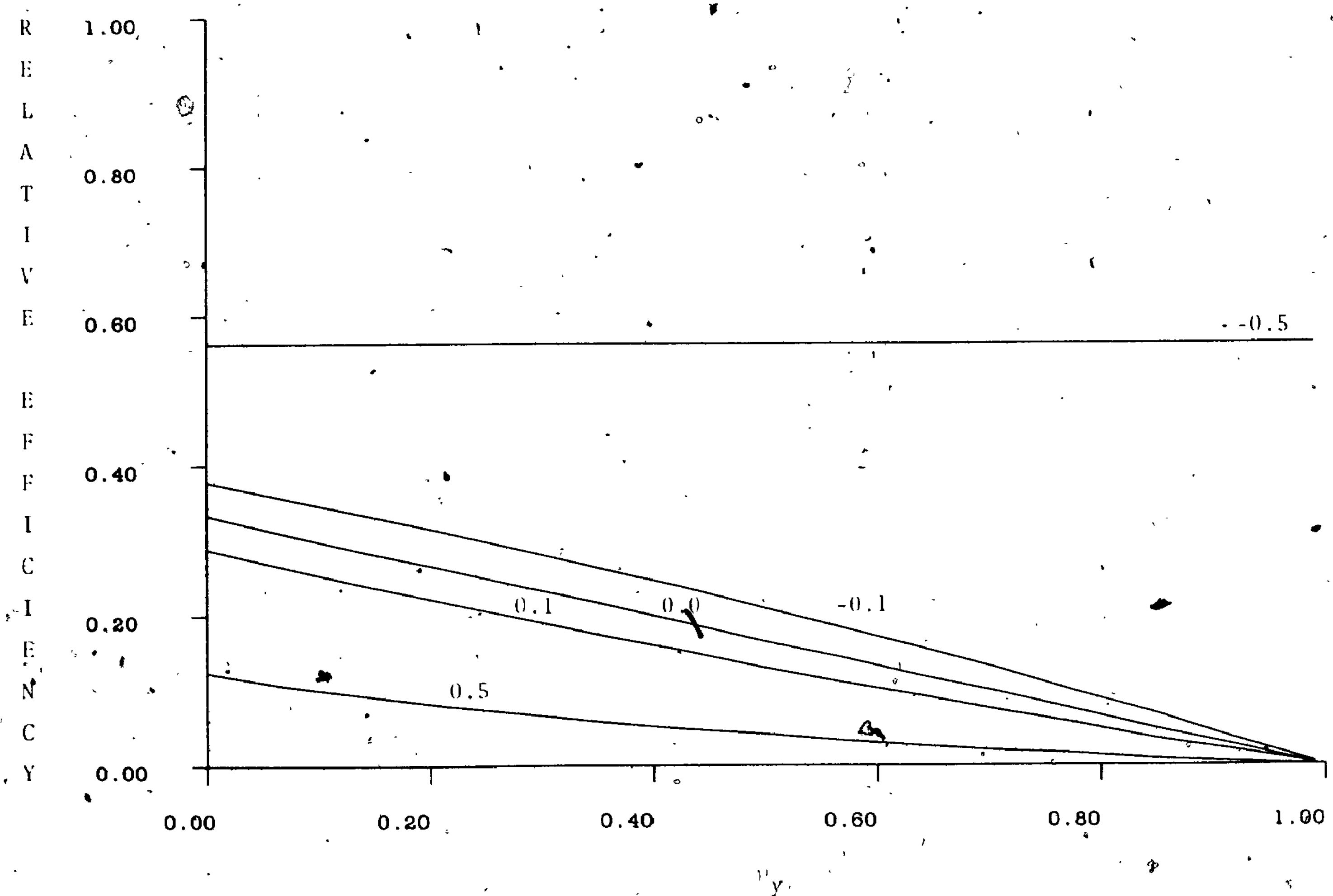


Figure 4.17: Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of ρ_x
(n=5, single binary covariate, β=0)

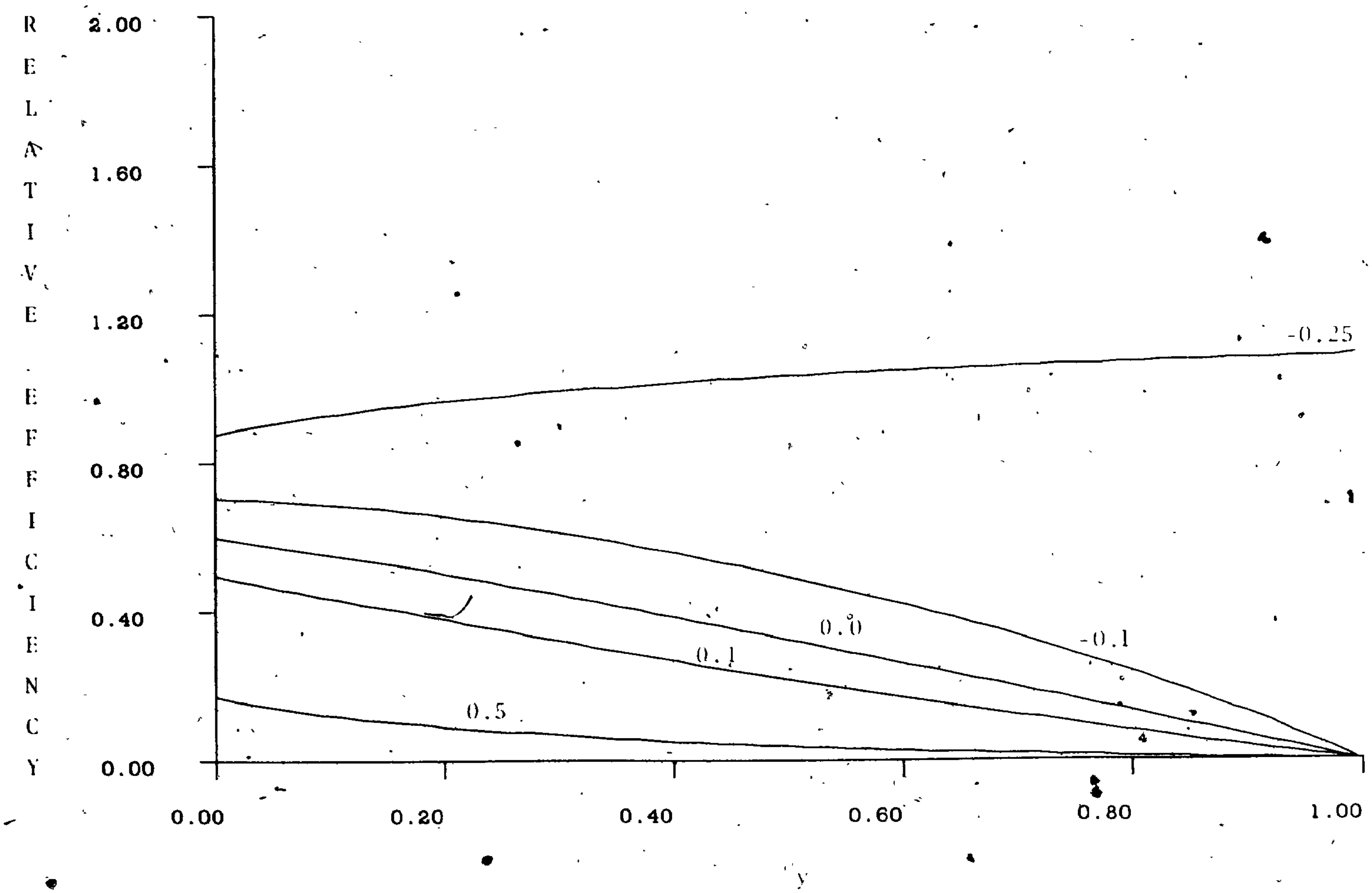


Figure 4.18. Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of λ
(n=10, single binary covariate, ρ=0)

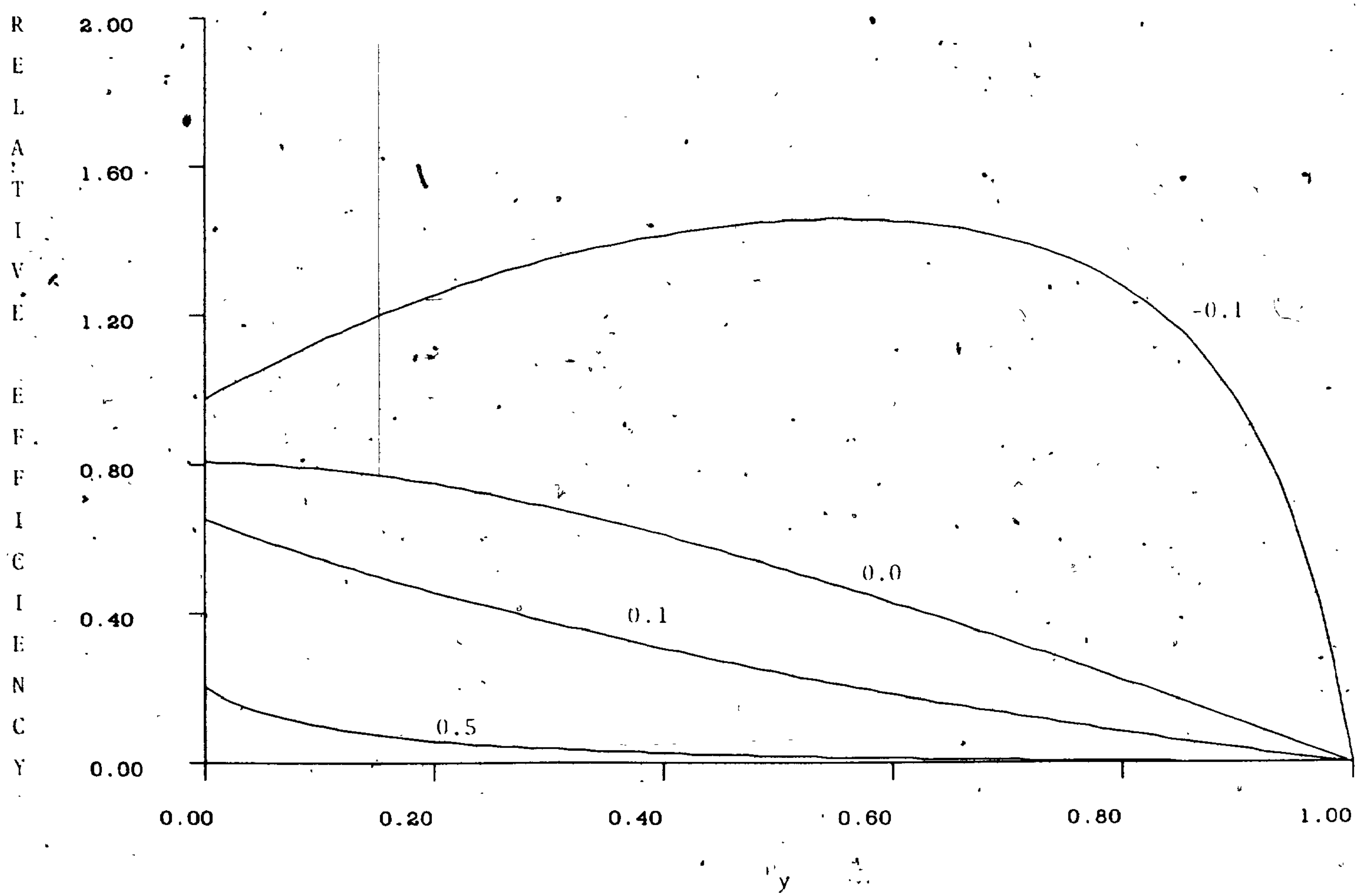


Figure 4.19. Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of λ (n=25, single binary covariate, ; =0)

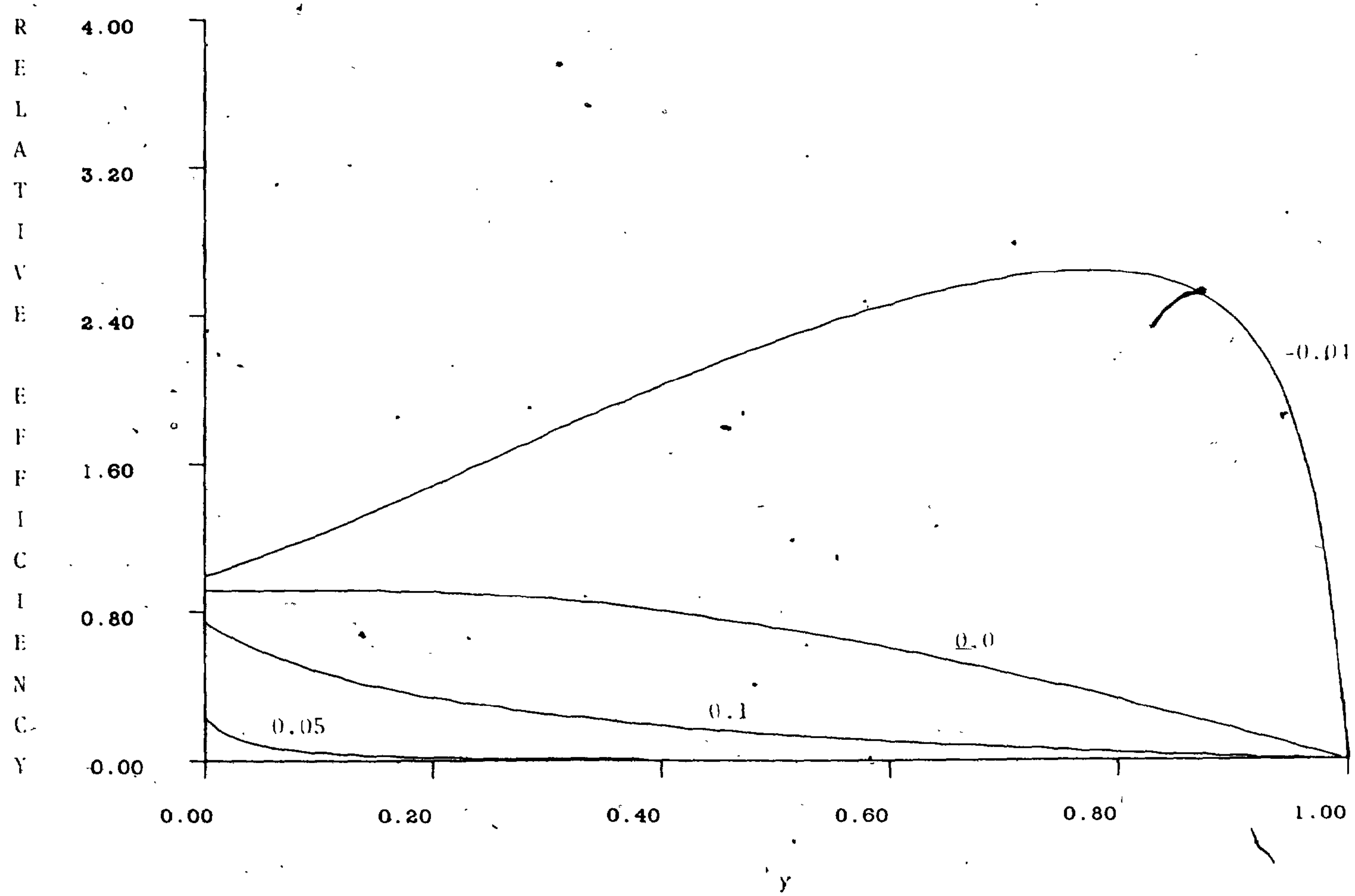
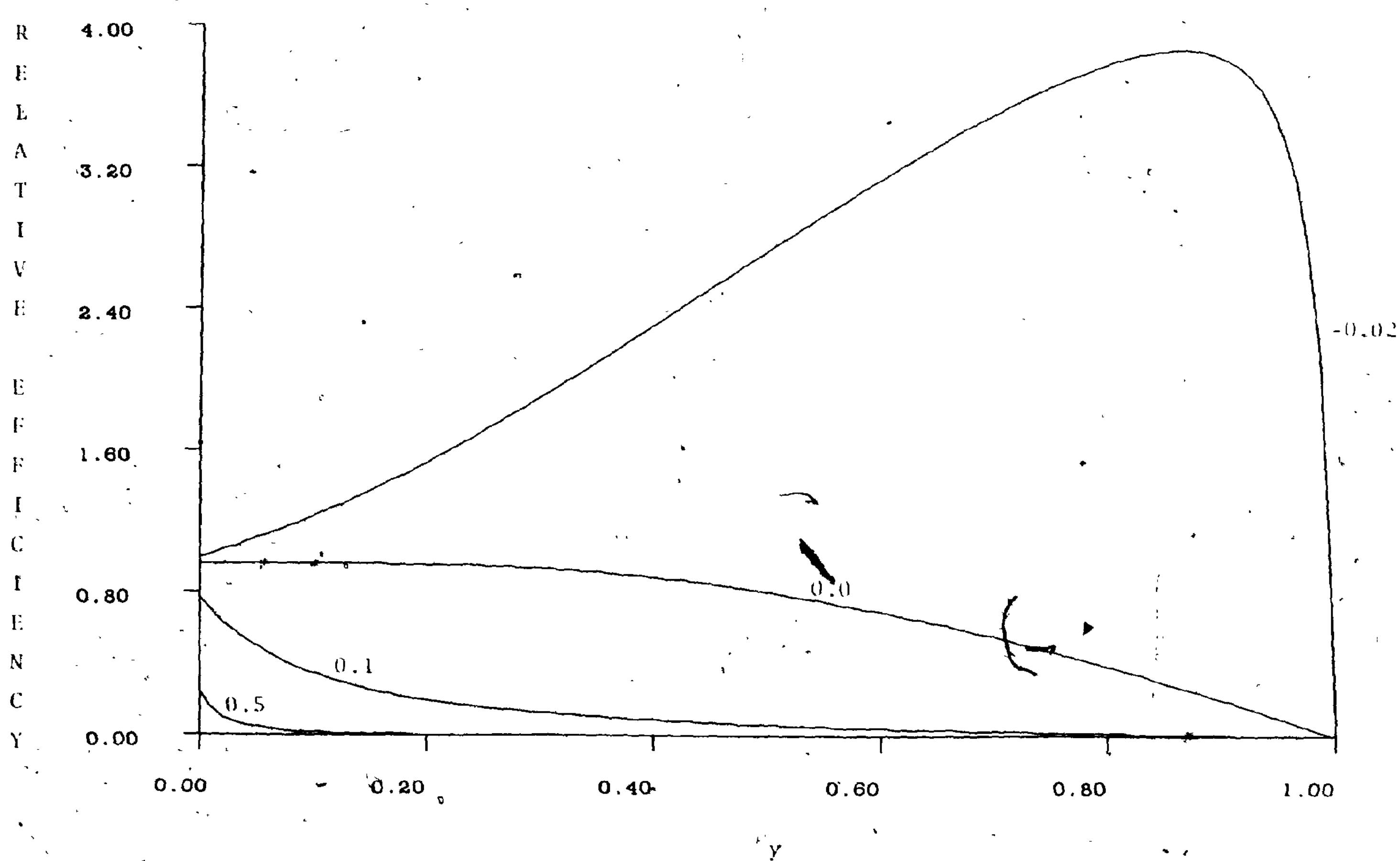


Figure 4.20. Asymptotic Relative Efficiency of the Dummy Variables Estimator for Varying Values of ρ_x
(n=50, single binary covariate, ρ=0)



$$= 4 / [\sigma_y^2 (1-\rho_y) \sigma_x^2 (1-\rho_x)]$$

which is four times the asymptotic variance of the conditonal estimator.

Now the same result follows from some work of Andersen(1973, p69) which was reported in Breslow and Day(1980, p.250). An alternative derivation using the notation common to this thesis is given in Appendix C.3. This result states that, in the case of a single binary covariate with two units per cluster, the dummy variables estimator is exactly twice the conditional estimator; hence the variance of the dummy variables estimator is four times that of the conditional estimator.

4.4.2 Non-zero regression coefficients

Since the usual and dummy variables estimators are, in general, asymptotically biased, and for small cluster size, the bias is quite large, it is of interest to compare only the consistent estimators of the regression coefficients, that is, the maximum likelihood (or unconditional) estimator with the conditional estimator.

In the preceeding section it was shown that, when

$$\beta = 0$$

the relative efficiency of the conditional estimator is never greater than one and is in general less than one, changing with increasing cluster size from very low values for cluster size two to larger values for ρ_y and ρ_x near zero, but to smaller values for ρ_y and ρ_x away from zero. In this section we investigate the relative efficiency of the conditional estimator for non-zero values of β in order to compare these values with those at

$$\beta = 0.$$

We limit our investigation to a single binary covariate with clusters of size 2.

We take the expectation of matrices calculated in Chapter 3 with respect to simple distributions of X and Y. For X we assume the simple correlated binary distribution given in Appendix A. This is actually a beta-binomial distribution with parameters 2, ρ_x and μ_x. we shall assume that

$$\mu_x = 0.5.$$

This corresponds to assuming that a and b, the usual parameters of the underlying beta distribution, are equal. The conditonal distribution of Y will be assumed to be Rosner's version of the correlated logistic. The parameter of this distribution is denoted by ρ_y.

For the unconditional estimator, minus the second derivative of the log likelihood is a simple form of expression (3.4.1.7), namely,

$$\begin{bmatrix} g_1(1-g_1) & -g_1g_2 & \begin{pmatrix} g_{11}-g_1f_1 & g_{12}-f_1f_2 \\ g_{21}-g_2f_1 & g_{22}-g_2f_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ -g_2g_1 & g_2(1-g_2) & \\ & & (x_1\ x_2)\begin{pmatrix} f_{11}-f_1^2 & f_{12}-f_1f_2 \\ f_{21}-f_2f_1 & f_{22}-f_2^2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{bmatrix}$$

where the submatrix in the lower left corner of this matrix has been omitted because it is merely the transpose of the submatrix in the upper right corner. This matrix may now be written

$$\begin{bmatrix} g_1(1-g_1) & -g_1g_2 & (g_{11}-g_1f_1)x_1+(g_{12}-f_1f_2)x_2 \\ & g_2(1-g_2) & (g_{21}-g_2f_1)x_1+(g_{22}-g_2f_2)x_2 \\ & & (f_{11}-f_1^2)x_{21}+2(f_{12}-f_1f_2)x_1x_2+(f_{22}-f_2^2)x_{22} \end{bmatrix} \quad (4.4.2.1)$$

where only the upper triangular submatrix is reported (because the matrix is symmetric) and where

$$\begin{aligned}
g_1 &= Pr(Y_1=1,Y_2=0)+Pr(Y_1=0,Y_2=1), \\
g_2 &= Pr(Y_1=1,Y_2=1), \\
f_1 &= Pr(Y_1=1), \\
f_2 &= Pr(Y_2=1), \\
g_{11} &= Pr(Y_1=1,Y_2=0), \\
g_{12} &= Pr(Y_1=1,Y_2=1), \\
g_{21} &= Pr(Y_1=1,Y_2=1) \\
g_{22} &= Pr(Y_1=1,Y_2=1) = g_2, \\
f_{11} &= Pr(Y_1=1) = f_1, \\
f_{12} &= Pr(Y_1=1,Y_2=1) = g_2, \\
f_{21} &= Pr(Y_1=1,Y_2=1) = g_2
\end{aligned}$$

and

$$f_{22} = Pr(Y_2=1) = f_2.$$

When one takes expectation of matrix (4.4.2.1) with respect to the distribution of (Y,X) one obtains a matrix, say A. The asymptotic variance of the unconditional maximum likelihood estimator is given by the inverse matrix, namely A^{-1}, and, in particular, the variance of $\hat{\beta}_R'$ is given by the element in position (3,3) of A^{-1}. If we define a_{ij} as the element in position (i,j) of the matrix A, then a result given by Graybill(1976, p.19) shows that the required element of A^{-1} is

$$\frac{a_{11}a_{22}-a_{12}^2}{a_{33}(a_{11}a_{22}-a_{12}^2) - (a_{22}a_{12}^2-2a_{12}a_{23}a_{13}+a_{11}a_{23}^2)} \qquad (4.4.2.2)$$

The asymptotic variance of the conditional estimator may be calculated in a similar way. A typical element in the matrix of (minus the second derivative of the log likelihood) is given by expression (3.4.3.3) and calculation of the unconditional information is shown by expression (3.4.3.8). In the case of one covariate and two units per cluster, the conditional information in a single cluster is

$$2E_X[X^2F(1-F) - X_1X_2F_1F_2] \qquad (4.4.2.3)$$

where F is the probability distribution function under the conditional model, that is

$$\begin{aligned} F &= Pr(Y_1=1) \\ &= \exp(\beta x_1)/[\exp(\beta x_1) + \exp(\beta x_2)] \end{aligned}$$

but this is only a function of x_1 and x_2, which we shall write as $f(x_1,x_2)$. Hence (4.4.2.3) becomes

$$\begin{aligned} 2\{&f(1,0)[1-f(1,0)]p(1,0)+f(1,1)[1-f(1,1)]p(1,1) \\ &- f(1,1)f(1,1)p(1,1)\} \end{aligned} \qquad (4.4.2.4)$$

where

$$p(i,j) = Pr(X_1=i\ X_2=j).$$

Now

$$f(1,0) = \exp(\beta)/[1+\exp(\beta)]$$

and

$$f(1,1) = 0.5,$$

so that (4.4.2.4) becomes

$$2 \exp(\beta)\ \sigma_x^2(1-\rho_x)\ /[1+\exp(\beta)]^2$$

where the distributon of X is the simple correlated binary (See Appendix A.1) To get the unconditional information we multiply this expression by g_1, the probability that one of the two Y's is 1. The asymptotic variance of $\hat{\beta}_c$ is the inverse of the resulting scalar, that is,

$$[1+\exp(\beta)]^2\ /[2g_1\ \exp(\beta)\ \sigma_x^2(1-\rho_x)]. \qquad (4.4.2.5)$$

Now when

$$\beta = 0,$$

g_1 is merely a function of the parameters of the conditional distribution of Y that is,

$$g_1 = 2\ \sigma_y^2(1-\rho_y)$$

so that the asymptotic variance of $\hat{\beta}_c$ is

$$[\sigma_y^2\ \sigma_x^2\ (1-\rho_x)\ (1-\rho_y)]^{-1}$$

as was given is Table 4.31.

In evaluating the relative efficiency of the conditional estimator we assume that the correlated binary distribution of X has

$$p_1 = p_0 = 0.5$$

so that

$$\sigma_x^2 = 0.25,$$

and that the underlying beta distribution of the error distribution of the Y´s has mean 0.5 so that

$$\sigma_y^2 = 0.25.$$

The subroutine RELEFF (see Appendix D.3) was written to evaluate the asymptotic variance of the unconditional and conditional estimators as given in expressions (4.4.2.2) and (4.4.2.5) and to calculate the relative efficiency of the conditiional estimator as the ratio of these two variances. The results were plotted as a function of ρ_y for

$$\exp(\beta) = 1,\ 2,\ 5,\ 10$$

(the inverse values 1, 0.5, 0.2, 0.1 giving the same values of the relative efficiency) and for

$$\rho_x = 0.5,\ 0.1,\ 0.0,\ -0.1,\ -0.5$$

(larger positive values giving even lower values of the relative efficiency).

The graphs in Figures 4.21 to 4.25 show that the relative efficiency of the conditional estimator decreases with

1. increasing values of ρ_y, the ´residual´ intraclass correlation,

2. increasing values of $|\beta|$, the absolute value of the regression coefficient,

3. increasing values of ρ_x, the intraclass correlation of the covariate

so that the maximum efficiency is attained for

$$\rho_y = 0.0,$$
$$\beta = 0.0,$$
$$\rho_x = -1.0.$$

It may be shown that the conditional estimator is as efficient as the unconditional estimator when ρ_x is -1.0, for all values of ρ_y and β.

The result in 2. may be interpreted in the following way:
The conditional estimator would use only those clusters with dissimilar values of y_1 and y_2. If the correlated logistic model is correct, the condtional estimator is fully efficient only if, within each cluster so used, the covariate values are dissimilar. The conditional estimator in the case of a single binary covariate with two units per cluster is discussed by Breslow and Day(1980, p.164-6), for use in case-control studies. If used in other types of studies where the correlated logistic provides a reasonable sampling model, the conditional estimator, which uses only those clusters with dissimilar x values, and ignores those clusters with similar x values, is usually not fully efficient. It disgards the information in the clusters with dissimilar y values but similar x values.

Figure 4.21. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of e^{β}

(π_x=0.5, single binary covariate, n=2)

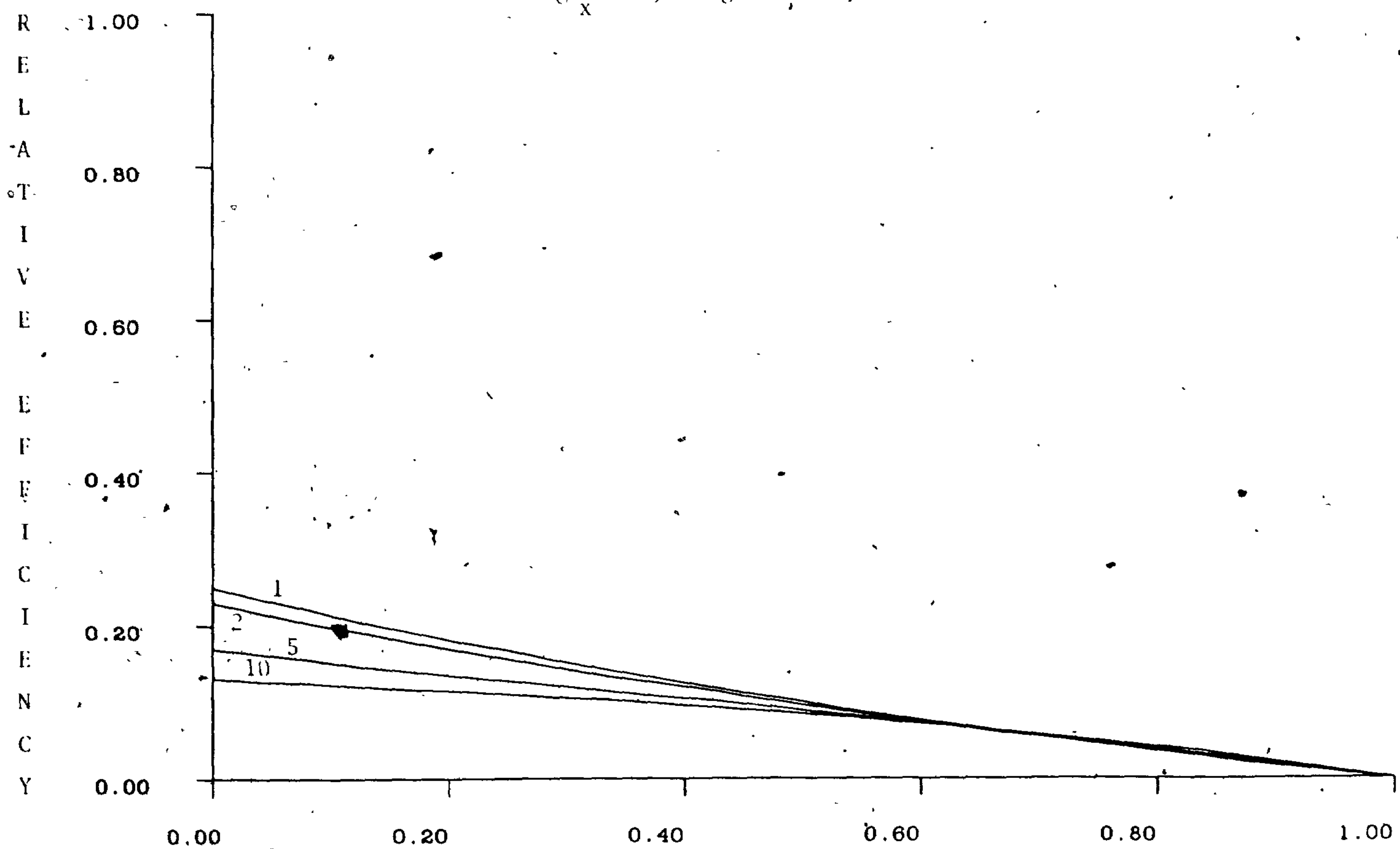


Figure 4.22. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of $e^{\prime}$

(θ_{x}=0.1, single binary covariate, n=2)

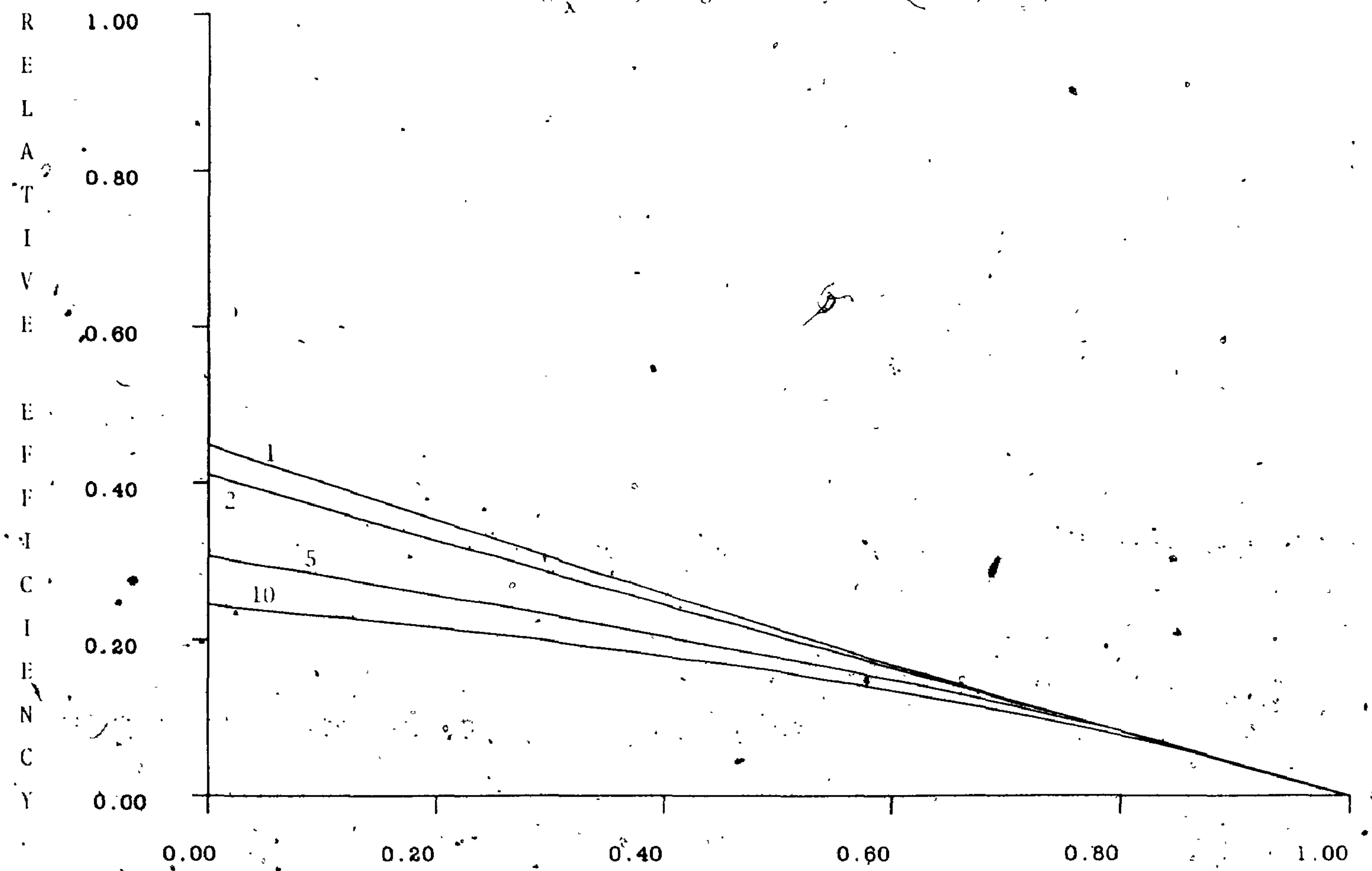


Figure 4.23. Asumptotic Relative Efficiency of the Conditional Estimator for Varying Values of [illegible]

(μ_x=0.0; single binary covariate, n=2)

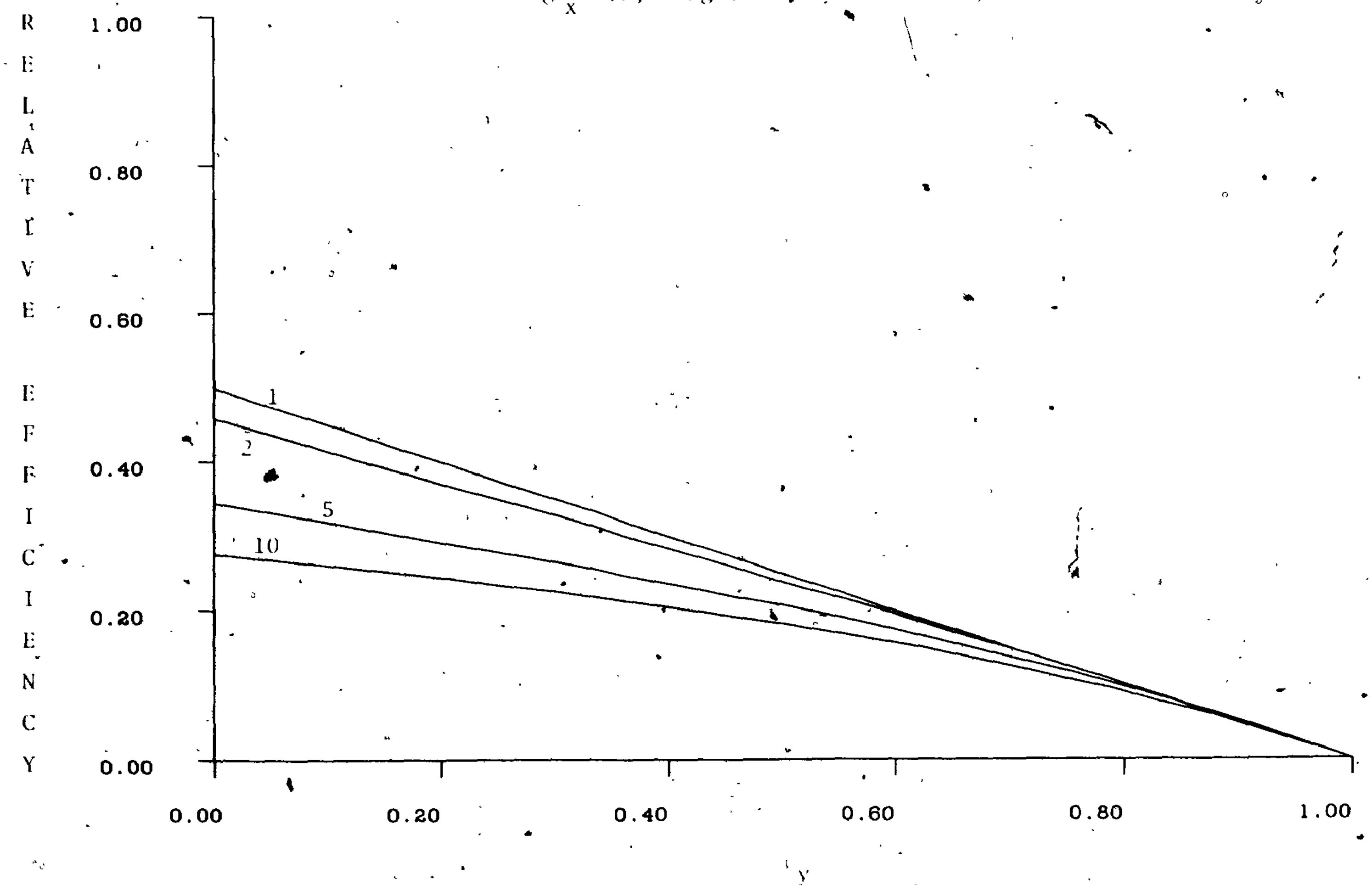


Figure 4.24. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of e^{β}

(${}_x$=-0.1, single binary covariate, n=2)

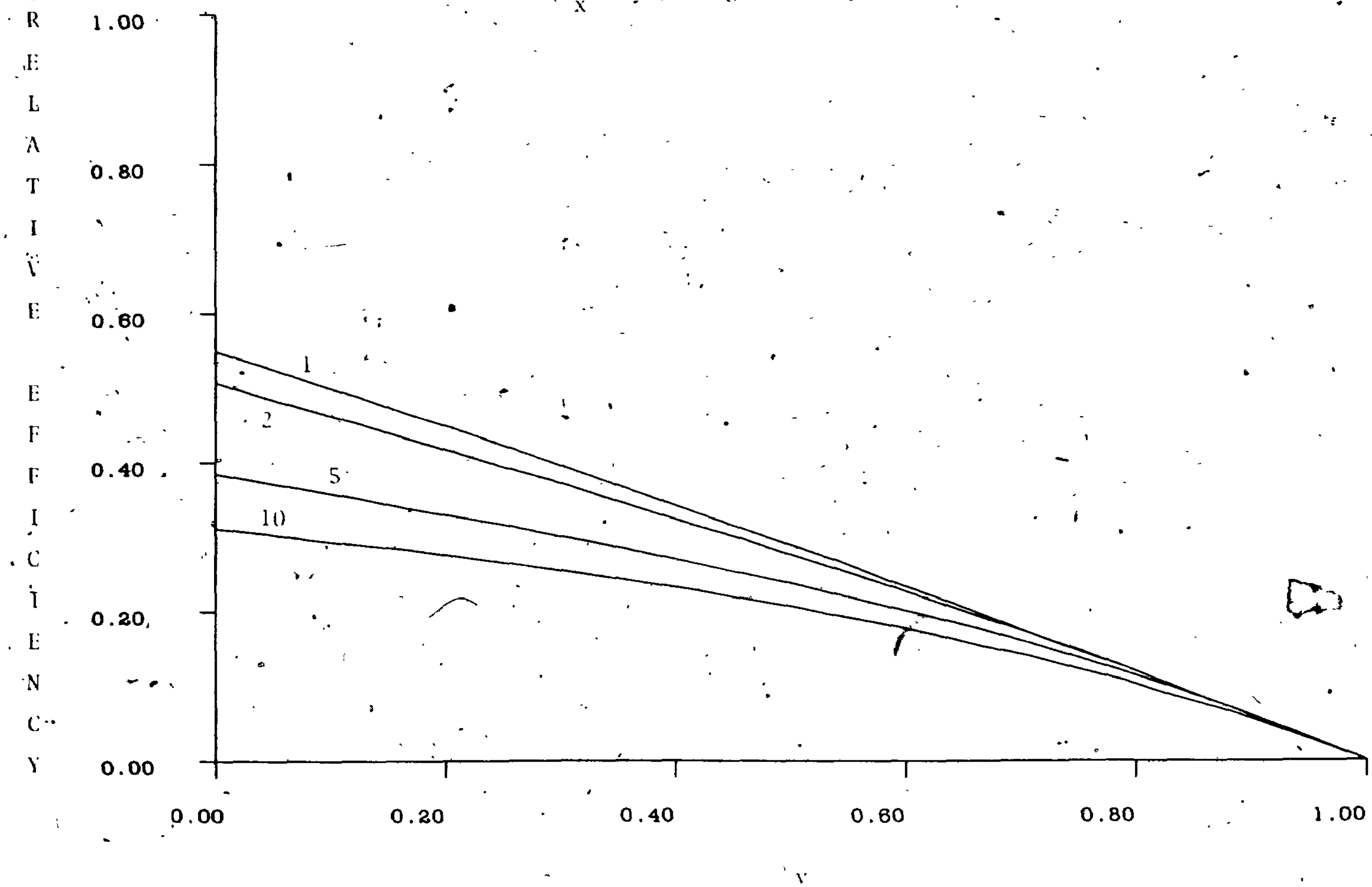
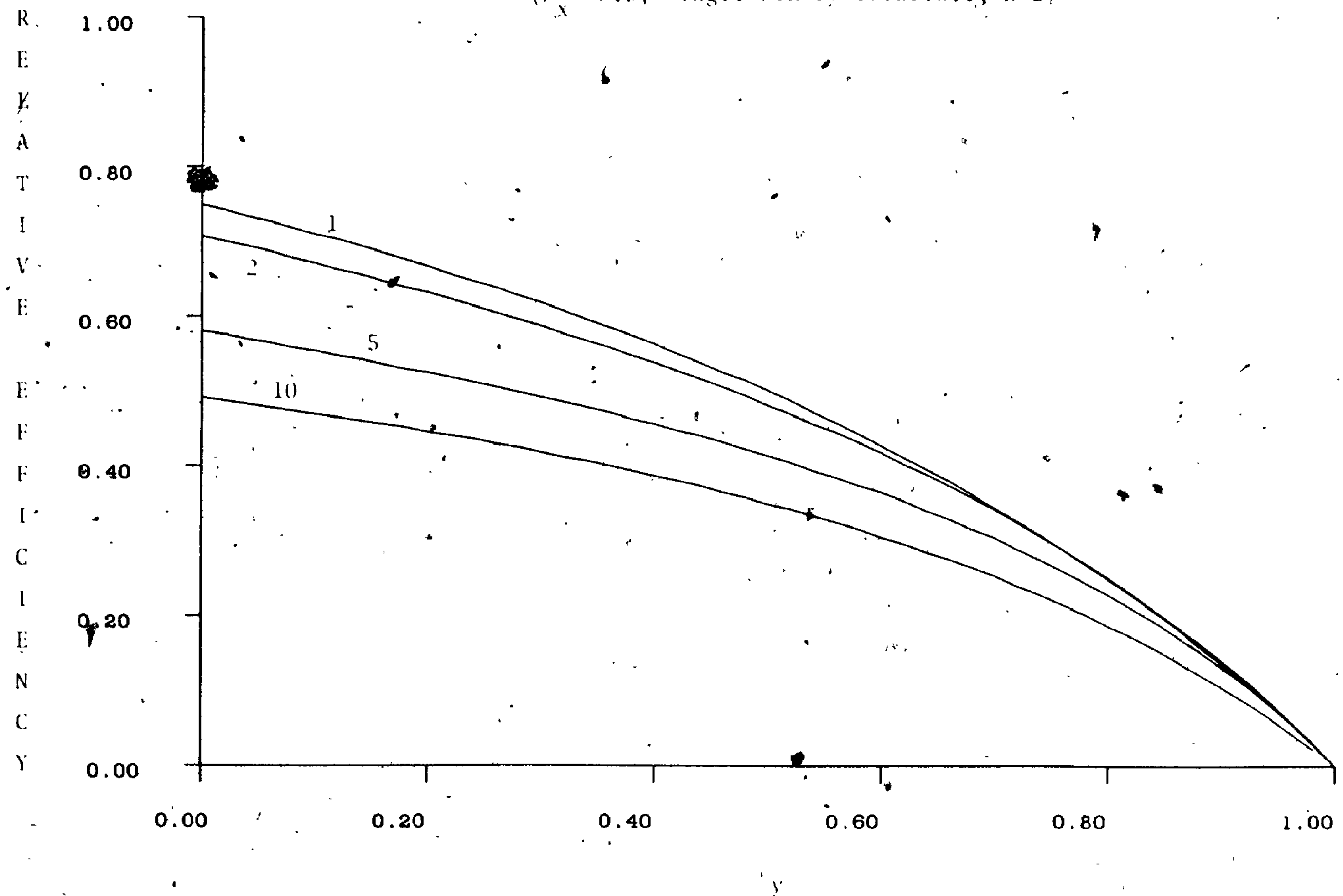


Figure 4.25. Asymptotic Relative Efficiency of the Conditional Estimator for Varying Values of c

(ρ_x=-0.5, single binary covariate, n=2)



If, however, we are actually dealing with a case-control study, in which there only exist clusters with dissimilar y values, then the unconditional model could not be used. There would be no clusters with 0 or 2 occurrences; these clusters must be present to enable us to estimate the parameter α of the unconditional model. The case-control study with a fixed number of cases per control is an example in which the correlated logistic does not provide a reasonable sampling model.

Hence the conditional estimator, although not fully efficient when all types of clusters may occur, is fully efficient in case-control studies.

4.5 <u>Misspecification factors</u>

The misspecification factor for the variance of an estimator is the amount by which the estimated variance of the estimator must be multiplied in order to produce an unbiased estimator of the true variance of that estimator. The misspecification is cause by the fact that the estimated variance is calculated under the wrong model, and the misspecification factor is used to adjust this estimate to produce a consistent estimate of the variance under the correct model. The misspecification effect was first described for continuous outcome variables by Scott and Holt(1982).

Both the usual estimator, $\hat{\beta}_u$, and the dummy variables estimator, $\hat{\beta}_d$, suffer misspecification effect. The misspecification factors for p covariates when

$$\boldsymbol{\beta} = \mathbf{0}$$

are

$$[E(\mathbf{XX}')+(n-1)\rho_y E(\mathbf{X}_1\mathbf{X}_2')]E(\mathbf{XX}')^{-1}$$
$$= I + (n-1)E(\mathbf{X}_1\mathbf{X}_2')E(\mathbf{XX}')^{-1}$$

and

$$nc[E(\mathbf{XX}')+(n-1)\rho_z E(\mathbf{X}_1\mathbf{X}_2')][E(\mathbf{XX}')-E(\mathbf{X}_1\mathbf{X}_2')]^{-1}/(n-1),$$

respectively. For a single covariate, these factors become

$$1 + (n-1)\rho_y\rho_x \qquad (4.5.1)$$

and

$$nc[1 + (n-1)\rho_z\rho_x]/\ [(n-1)(1-\rho_x)] \qquad (4.5.2).$$

Expression (4.5.1) is exactly that obtained by Campbell(1979) and Scott and Holt for the continuous case. Note that, for ρ_x values greater than zero,the factor is greater than 1 and multiplication of the estimated variance will yield a larger standard error; thus the usual test is too optimistic. However, for ρ_x less than 0, and in particular for designed studies, the factor is less than 1 and the new variance will be smaller than before adjustment; hence the usual test is conservative.

The values of the misspecification factor for the dummy variables variance given by (4.5.2) varies quite widely as indicated by the plots in figures 4.26 to 4.31. It can be seen that

1. the misspecification factor is, in general, greater than 1, indicating that the estimator, even when unbiased, provides overly optimistic tests.

2. the misspecification factor decreases with increasing values of n, after starting at some very large values for

$$n = 2.$$

3. the misspecification factor increases with increasing values of ρ_x and, in general, with increasing values of ρ_y.

4. starting with moderate values of n, say

$$n = 10,$$

there are some negative values of ρ_x and values of ρ_y, for example,

$$\rho_y < 0.40,$$

for which the misspecification factor is essentially 1, that is, for which no correction is required in the estimation of the variance of the dummy variables estimator.

Figure 4.26. Misspecification Factor for the Dummy Variables Estimator
(n=2, single binary covariate, ρ=0)

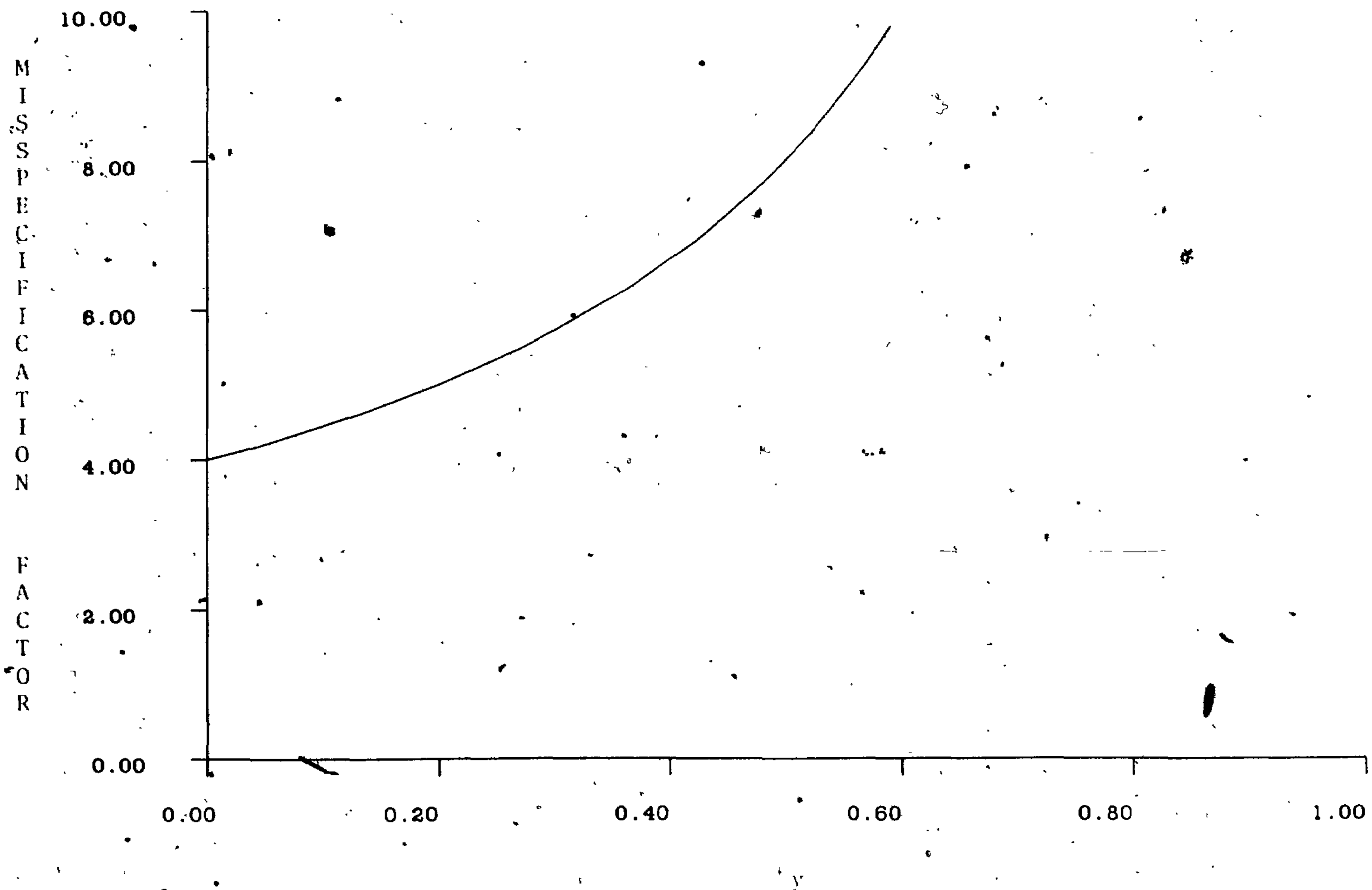


Figure 4.27. Misspecification Factor for the Dummy Variables Estimator for Varying Values of λ (n=5, single binary covariate, $\rho=0$)

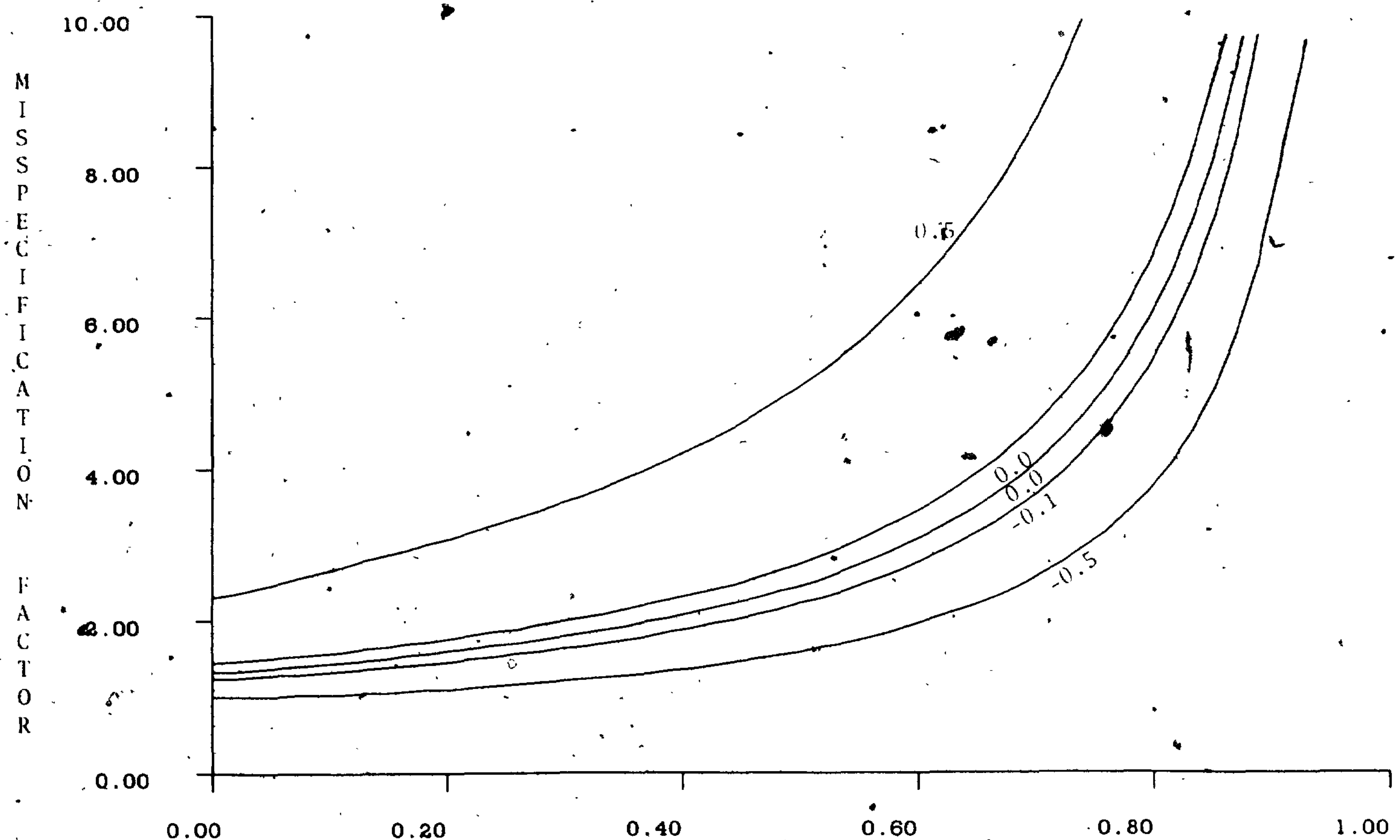


Figure 4.28. Misspecification Factor for the Dummy Variables Estimator for Varying Values of λ
($n=10$, single binary covariate, $\rho=0$)

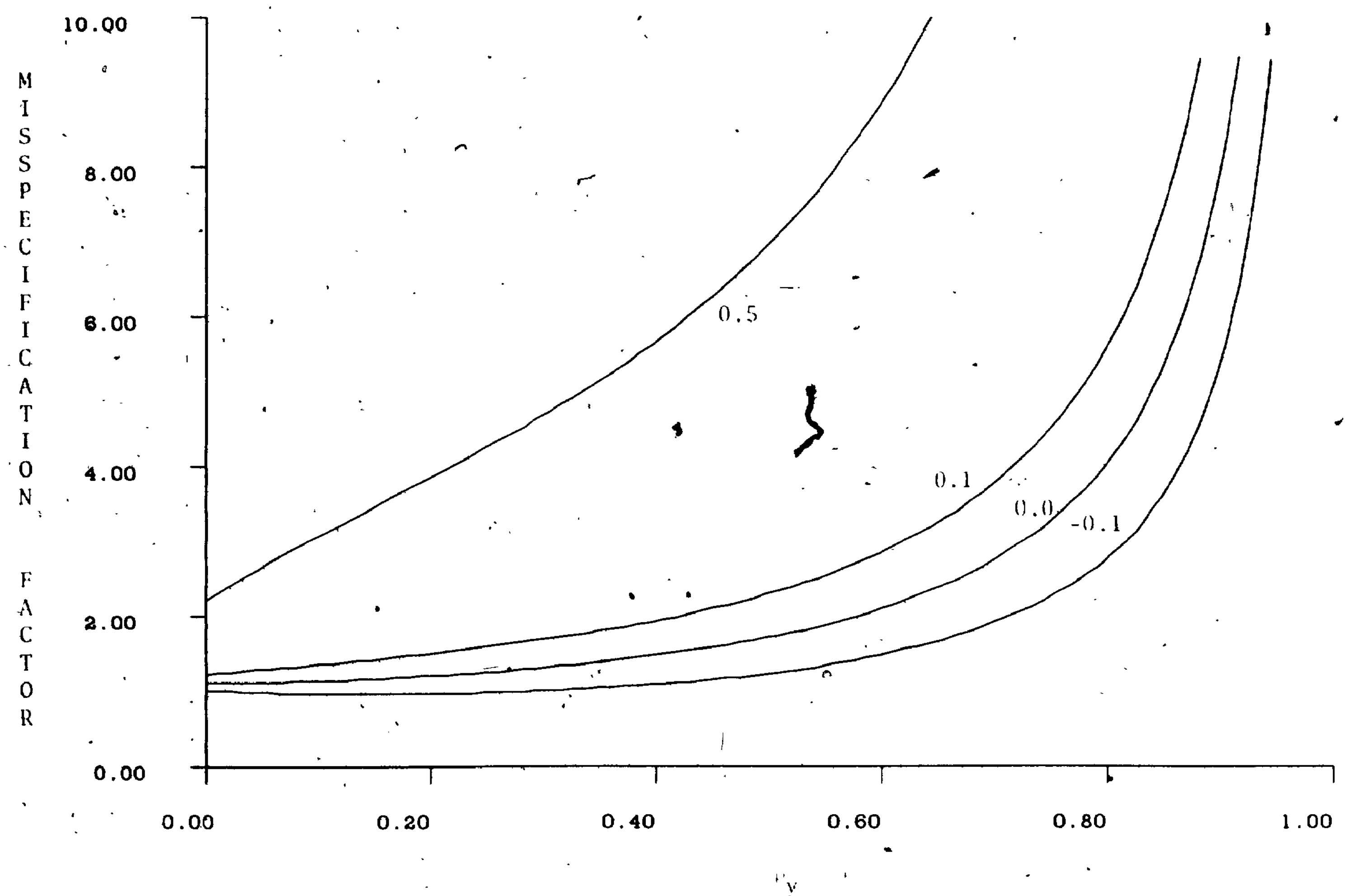
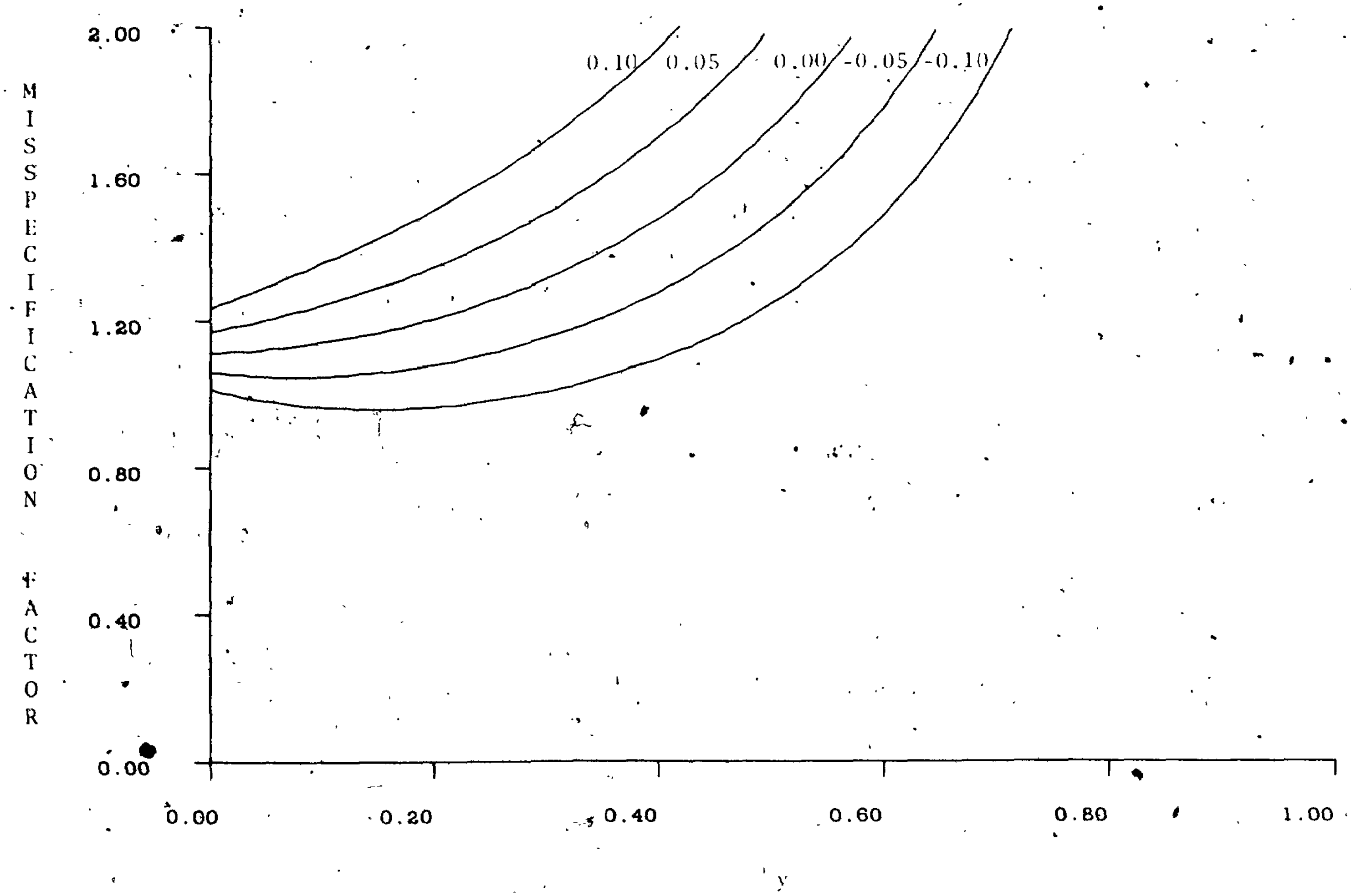


Figure 4.29. Misspecification Factor for the Dummy Variables Estimator for Varying Values of ρ_x

(n=10, single binary covariate, $\rho=0$)



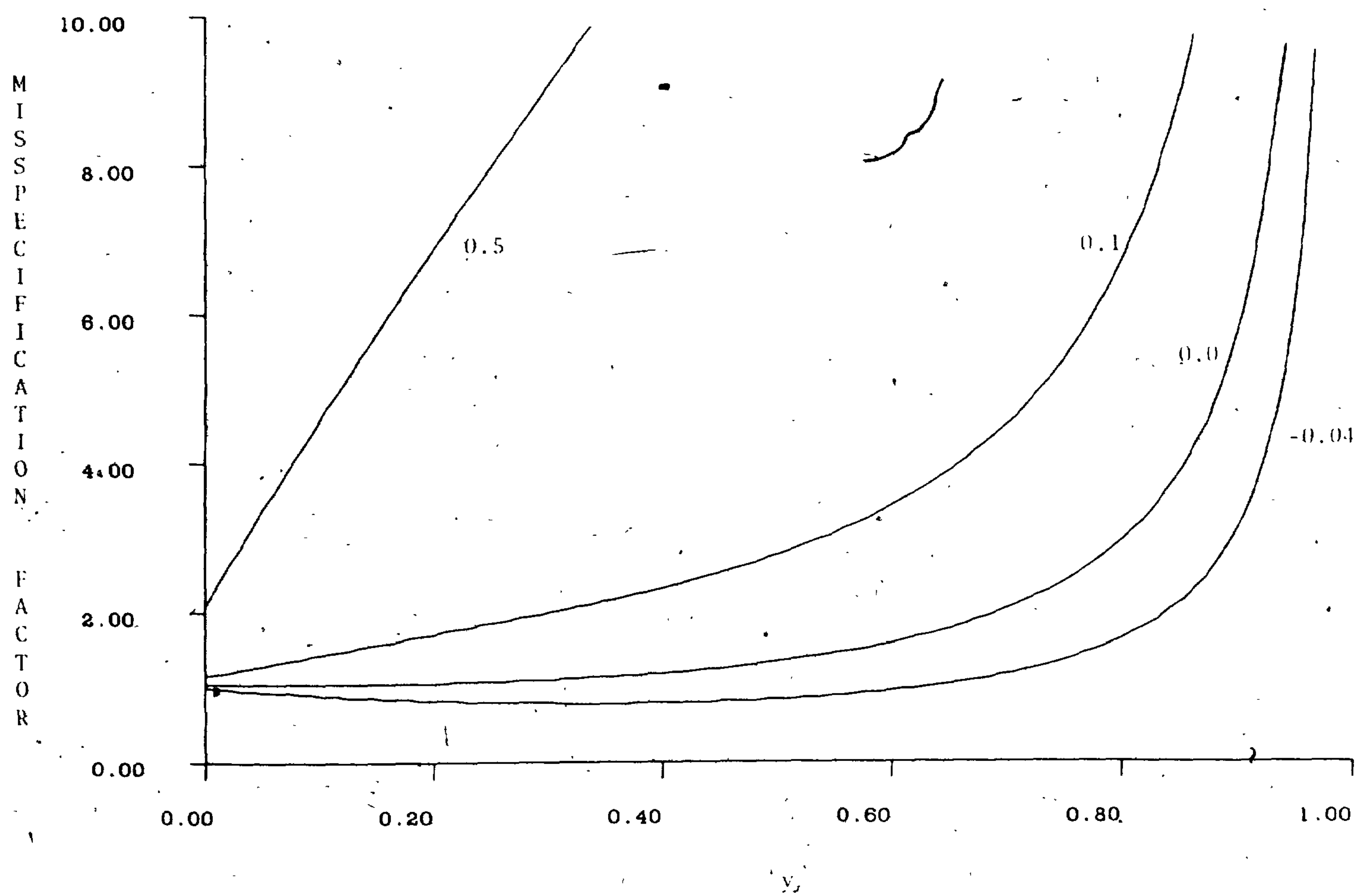
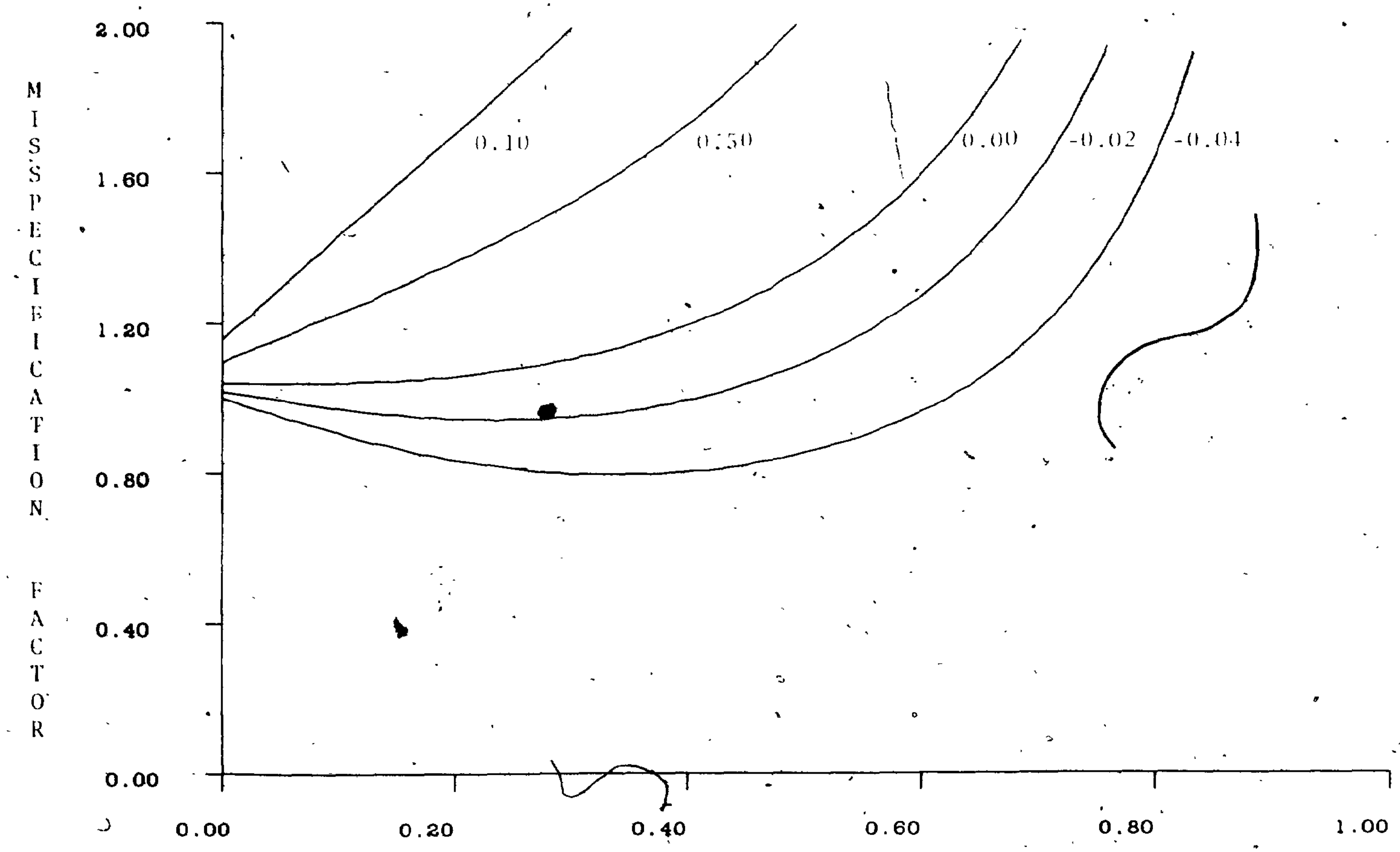


Figure 4.30. Misspecification Factor for the Dummy Variables Estimator for Varying Values of x

(n=25, single binary covariate, r=0)

Figure 4.31. Misspecification Factor for the Dummy Variables Estimator for Varying Values of ρ_x
(n=25, single binary covariate, ρ=0)



CHAPTER 5

Numerical comparisons with sample data

5.1 Introduction

In this chapter we perform some numerical calculations with the various estimators, obtaining their values in two examples, as well as estimates of their variances and/or equivalent test statistics, and show that, as the results of Chapters 3 and 4 suggest, numerical values of the usual, dummy variables and conditional estimators may differ greatly from those of the unconditional maximum likelihood estimator, and may produce test statistics which may give inferences quite different from that provided by the unconditional approach. In this chapter we use Rosner's version of the correlated logistic model.

Note that both examples have only unit-specific covariates. Although the usual and maximum likelihood estimators may be used with either cluster- and/or unit-specific covariates, the conditional estimator may only be used with unit-specific covariates (see Breslow and Day(1980)). There are some numerical problems with the calculation of the dummy variables estimator when only cluster-specific covariates are used. Because the dummy variables are also cluster-specific, there is confounding of the effects of the dummy variables with those of the covariates. The standard logistic regression analysis program will fail due to singularity of the matrix of second derivatives. However, reparametrization and the removal of some variables from the model will allow a numerical solution and the correct calculation of estimates.

The computer programs used for these calculations were

1. LOGDIS, which implements a version of J.A.Anderson's algorithm for the multinomial (polychotomous) logistic model. Documentation is contained in Koval(1982). This program was used to calculate both the usual and dummy variables estimators, their estimated variances, and test statistics (both estimate/(estimated standard error) and likelihood ratio test).

2. CLOG, a version of the conditional likelihood program for case-control studies used by Breslow and Day(1980, p.307-21). This program provides numerical values of the conditional estimator, its estimated variance, and the likelihood ratio test.

3. CORLOG, written by the author to calculate the maximum likelihood estimator under Rosner's constrained version of the correlated logistic model. However the calculation of the elements of the matrix of second derivatives as described in expression (3.4.1.7)

with the modifications given at the end of section (3.4.2) is complicated by the need to determine the derivatives of the a parameters with respect to the two 'Rosner' parameters for each sample size. This program substitutes a likelihood ratio test for the test based on the ratio (estimate/(estimated standard error)).

5.2 Example one

This example has been artificially constructed as the simplest type of data that could be handled by all four methods. The simplicity of the example enables us to see the behaviour of the estimators quite clearly. The data is given in Table 5.1. It is divided into two parts, the first having data with dissimilar values of y, which can be analysed by the dummy variables and conditional estimation approach, the second containing results for those clusters with similar values of y, which are used only in the construction of the usual and the maximum likelihood (or unconditional) estimators. The results of the analysis using all four estimators is given in Table 5.2.

Table 5.1

Data for example one

Cluster	value of x	value of y
1	1,0	1,0
2	1,0	1,0
3	1,0	0,1
4	1,1	1,0
5	0,0	0,0
6	0,0	0,0
7	1,0	0,0
8	1,0	0,0
9	1,0	1,1
10	1,0	1,1
11	1,1	1,1
12	1,1	1,1

Table 5.2

Results of analysis of example one

Estimator	value	value/se	LR test
maximum likelihood	1.96		4.67
conditional	0.69	0.57	0.34
usual	1.79	1.98	4.33
dummy variables	1.39	0.80	0.68

5.3 Example two: Smith and Pike data

The second example consists of data on the age and presence of infection with Trypanosoma Cruzi in 40 households in a rural area in Brazil. These data were first presented by Mott et al.(1976) and then

Table 5.3

Example two:

Age and presence of *T.Cruzi* in 40 households in Brazil

(age followed by number of individuals at that age in household; number of infectives shown in parenthesis)

cluster	age,number(infectives)	total
1	1-1,5-1,10-4(2),25-1,65-1(1)	8(3)
2	10-1(1),45-1(1),65-1	3(2)
3	1-2,25-1,45-1	4(0)
4	5-1,10-1,35-1(1)	3(1)
5	25-1,55-1	2(0)
6	1-2,5-1,20-1(1),35-1	5(1)
7	5-1,15-1,25-1	3(0)
8	20-1,55-2(1)	3(1)
9	0-1,1-1,5-1,10-1,25-1(1),35-1(1)	6(2)
10	15-1,55-2	3(0)
11	5-1,45-1,55-1	3(0)
12	5-1,20-1(1),25-1	3(1)
13	1-1,25-2(2)	3(2)
14	45-1	1(0)
15	5-2(2),15-1(1),20-1(1),35-1(1),45-1(1)	6(6)
16	5-2,20-1(1),25-1(1)	4(2)
17	1-1,20-1(1),25-1	3(1)
18	0-1,1-1,25-1	3(0)
19	15-1(1),25-1(1)	2(2)
20	1-2,5-2,10-3,15-1,20-2,35-1,45-1(1),65-1(1)	13(2)
21	45-1(1),55-1	2(1)
22	1-1,5-2,10-1,25-1	5(0)
23	5-4,10-1,25-2(1),35-1,55-1	9(1)
24	5-1,10-1(1),15-1,25-1(1),35-1(1)	5(3)
25	20-1,25-1,55-1,65-1	4(0)
26	5-1(1),15-1(1),25-2(1),65-2(1)	6(4)
27	1-1(1),5-2(2),25-2(2)	5(5)
28	0-1,20-1,25-1,65-2	5(0)
29	65-2(2)	2(2)
30	15-1,45-1,65-1	3(0)
31	15-1(1),25-2	3(1)
32	45-2(1)	2(1)
33	25-2[illegible])	2(1)
34	5-1(1),10-1(1),65-1	3(2)
35	1-2,20-1,35-1(1)	4(1)
36	5-1,10-1(1),20-1(1),25-1(1),35-1(1),55-1(1)	6(5)
37	15-1,25-1(1)	2(1)
38	1-2,25-1(1),35-(1)	4(2)
39	0-1,5-2,10-1,35-2(1),65-1	7(1)
40	20-2	2(0)

reported by Smith and Pike(1976). They are given in Table 5.3 in a format suitable for computer entry and analysis. The age given in the table is the lower boundary of the age grouping as described by Pike and Smith. They do not use the age data explicitly, but we require a measurement of the age covariate for each unit. We have used this lower boundary of the age grouping as the age of the individual. The results of the analysis is given in Table 5.4.

Table 5.4

Results of analysis of example two

Estimator	value	value/se	LR test
maximum likelihood	0.0139		2.80
conditional	0.0358	2.66	7.91
usual	0.0167	2.01	4.07
dummy variables	0.0456	2.94	9.99

5.4 Conclusions

From the analyses of these two data sets, the following conclusions may be drawn

1. any estimator which ignores the clusters with similar values of the outcome variable is at a disadvantage. In these examples, the large number of such clusters indicate a high degree of

intracluster correlation, but the correlation is not measured by estimators, such as the conditional and dummy variables, which discard these clusters.

2. the usual estimator performs quite well. In example one, it is slightly smaller than the unconditional estimator. The results of Chapter 4 would suggest that this is due to the negative intraclass correlation of the covariate (estimated at -0.134). In example 2, the usual estimator is larger than the unconditional (Rosner) estimator. Again Section 4.2.1 would suggest that this is due to the positive intraclass correlation of the covariate (estimated at 0.109) and the significant residual intracluster correlation (estimated at 0.192, LR test value of 12.14).

CHAPTER 6

Conclusions and areas of further study

This thesis has introduced the correlated logistic regression model, established some of its properties, and quantified the errors introduced by using incorrect models to estimate parameters of the correlated logistic model.

In Chapters 1 and 2 we have shown the relationship between the correlated logistic and other models for correlated binary outcomes. We have determined that it is a simplification of a model introduced by Cox(1972). Conversely, it is a generalisation of several well-known models, the beta-binomial of Skellam(1948) and Kleinman(1973), the beta-binomial with covariates of Williams(1982) and Manly(1978), the binary logistic of Berkson(1944,etc.), Day and Kerridge(1967), the multinomial logistic of Cox(1966) and Mantel(1966), and the simple correlated logistic of Rosner(1983). We have compared its properties with other models for binary outcomes in the presence of covariates, such as the logistic-normal of Pierce and Sands(1975) and Stiratelli, Laird and Ware(1984), and the equicorrelated probit of Ochi and Prentice(1984). We have distinguished it from the multivariate logistic of Mantel(1966) and Nerlove and Press(1973). We have shown that it can be used to produce the same conditional model used in case-control studies by Breslow, Day et al.(1978).

In chapter 3 we have investigated the consistency and asymptotic variance of several estimators of the regression coefficient of the correlated logistic model, in particular, the unconditional maximum likelihood, the conditional maximum likelihood, the usual (uncorrelated) logistic and a 'dummy variables' logistic estimator. We have established the regularity conditions for the consistency and asymptotic normality of the unconditional and conditional estimators, and have obtained sufficient conditions for the consistency of both the usual and dummy variables estimators. The sufficient conditions are:

1. for the complete set of p regression parameters,

$$\boldsymbol{\beta} = 0,$$

2. for a subset, $\boldsymbol{\beta}_a$, of the parameters,

$$\boldsymbol{\beta}_a = 0$$

and $\mathbf{X}_a$ is distributed independently of $\mathbf{X}_b$, where $\mathbf{X}_a$ is the subset of the covariates corresponding to the subset $\boldsymbol{\beta}_a$ of the parameters, and $\mathbf{X}_b$ is the remainder of the covariates in $\mathbf{X}$.

We have obtained the asymptotic variance of the unconditional maximum likelihood estimator in the general case, when

$$\boldsymbol{\beta} = \mathbf{0},$$

and when Rosner's constrained version of the model is used. We have obtained the asymptotic variance of the conditional estimator in the general case and when

$$\boldsymbol{\beta} = \mathbf{0}.$$

Finally we have obtained the asymptotic variance of the usual and dummy variables estimators, and hence their asymptotic relative efficiency and misspecification factor.

In Chapter 4 we have investigated the relationship between the estimators for some simple cases and have shown

1. for a single binary covariate, the bias of the usual estimator increases as the cluster size increases. Hence the usual estimator is only useful for estimation and testing when

   $$\beta = 0.$$

   It would be interesting to investigate if this result holds in the presence of multiple covariates.

2. for the usual estimator with two correlated binary outcomes in clusters of size 2, the effect of the covariate with coefficient equal to 0 in never overestimated, and the effect of the covariate with non-zero coefficient is estimated with the same bias as in the single covariate case, although the size of the bias is somewhat smaller. It would be useful to investigate the effect of a non-zero parameter on the estimation of another non-zero parameter, and to study this further for multiple covariates.

3. for a single binary covariate, the bias of the dummy variables estimator becomes smaller with increasing cluster size, becoming negligible (about 5 percent) when n is 25. This indicates that the dummy variables estimator may be used for estimation in larger clusters, in agreement with the conclusion of Smith, Pike and Hill(1980). It would be of interest to investigate if these results hold for other types of covariates and for multiple covariates.

4. for a single binary covariate, the relative efficiency of the conditional estimator is always less than one, except in the case

$$\beta = 0$$

and

$$\rho_{x} = -1/(n-1),$$

the latter indicating a 'designed' study, for example, one in which a subject undergoes two treatment conditions, and the covariates measures the presence or absence of a treatment.

Moreover, the relative efficiency for non-zero ρ_y and non-zero ρ_x reaches 0.9 only for clusters of size 25 or more. Hence the conditional estimator should only be used with 'designed' studies (for example, case-control studies with fixed number of controls for each case) or for large cluster size and small values of ρ_x and ρ_y. Moreover, these results indicate that, if there are varying numbers of controls for each case in a case-control study, the conditional method is probably not fully efficient. This should be further clarified by some calculations with varying cluster size. This lack of high efficiency of the conditional estimator as compared to an unconditional estimator differs from the results of Farewell(1979) and Lubin(1981) (who did, hovever, have a different unconditional model from that used in this thesis).

5. for a single binary covariate in clusters of size 2, the relative efficiency of the conditional estimator decreases with increasing values of the absolute value of β. This means that the conclusions about poor efficiency of the conditional estimator described in 4. also hold for non-zero β. This result should be confirmed for larger cluster sizes and multiple covariates.

6. for a single binary covariate, the relative efficiency of the usual estimator is greater than 1.0 for negative values of ρ_x, in particular, for designed studies. However, the omnipresence of bias still precludes any recommendation for this estimator.

Studies using a mean square measure of efficiency would determine if there is any usefulness for the usual estimator. Note that the misspecification factor for this binary outcome study is the same function as that obtained in the continuous outcome case by Scott and Holt(1982).

7. for a single binary outcome the relative efficiency of the dummy variables estimator is very low for clusters of size 2, although it does increase with cluster size, becoming greater than 1.0 for some negative values of ρ_x. However, the relative efficiency for small non-negative values of ρ_x becomes greater than 0.9 only for clusters of size greater than 10 and small values of ρ_y. Hence, from the results in 3. and here, the dummy variables estimator could only be recommended for clusters of size greater than 25 with small values of the intraclass correlations, ρ_x and ρ_y.

From these results it can be seen that none of the possible alternative estimators have both consistency and high efficiency, so that, despite the higher computing costs of fitting the full unconditional correlated logistic regression model, the maximum likelihood estimator should be used.

The examples in Chapter 5 show that, in a typical data analysis, use of the wrong estimator, whether the conditional, the usual or the dummy variables, can lead to estimates quite different from the unconditional maximum likelihood estimate, as well as different inferences, sometimes suggesting that the regression coefficient is non-significant, when the unconditional estimator indicates that it is significant (see example one), or that it is significant when the unconditional test shows that it is not (see example two).

Finally there is a need for the development and interpretation of the correlated logistic regression model. Although it has been used by Rosner in its simplest form for the study of eye disease, it has not yet been used in a practical study for cluster size larger than 2, or for varying cluster size. The example in Chapter 5 shows that such model fitting is possible.

However, there is a need for the interpretation of the parameters of the more complicated models, that is, those with n α parameters $\{\alpha_i, i=1,\ldots n\}$, and for the introduction of easily interpretable models with fewer α parameters, such as that provided by Rosner. For example, one might use the two-parameter correlated binomial of Kupper and Haseman(1978), or the three-parameter correlated binomial of Paul(1984), or some other simplification of the correlated binary model of Bahadur(1961).

Moreover the correlated logistic model needs to be expanded to allow the correlation between two units to vary according to some class to which the units belong; for example, for family studies, we would like α_2 of the current correlated model to be written as α_{11}, if both units are male sibs, α_{12} if they are of opposite sex, and α_{22} if they are both females. It would, however, be necessary to determine how the other α parameters would be written under this new model. Nevertheless, this would provide a model for binary outcomes similar to that for continuous outcomes described by Donner and Koval(1981). Another model would allow α_2 to change according to the 'distance' between the two units; for example, assuming that the outcomes were repeated mesurements over time, we might be able to create a model for binary outcomes corresponding to the models with autocorrelated error structure for continuous outcomes.

In summary, we have described a new model for correlated binary observations in the presence of covariates, and have shown that other estimators, which fail to model the intraclass correlation explicitly are, in general, not consistent , and/or have low asymptotic relative efficiency, although there are a few special cases when one or other of these estimators perform as well as or better than the unconditional maximum likelihood estimator. It is conjectured that, in the presence of more covariates, these alternative estimators would perform even worse than indicated in this thesis. We have provided a model for analysing data which occurs fairly commonly, whether in family studies, repeated measures studies, or in the analysis of measurements on eyes, fingers, and other highly correlated experimental or observational units. We have developed computer software for handling some of these analyses. This software can be further modified to accomodate the full correlated logistic regression model.

Appendix A

Derivation of a correlated binary distribution.

Assume that the marginal distributions of X_1 and X_2 are identical, but that X_1 and X_2 are correlated. Hence let

$$p_1 = Pr(X_i=1),\ i=1,2,$$

$$p_0 = Pr(X_i=0) = 1 - p_1,\ i=1,2$$

$$p_{jk} = Pr(X_1=j, X_2=k),\ j,k=0,1,$$

$$p_{k|j} = Pr(X_2=k|X_1=j),\ j,k=0,1,$$

and

$$\rho = corr(X_1, X_2).$$

Hence ρ may be written

$$cov(X_1,X_2)/Var(X_1) = \{E(X_1X_2) - [E(X_1)]^2\}/\{E(X_1^2)-[E(X_1)]^2\}$$

$$= (p_{11} - p_1^2)/(p_1 - p_1^2).$$

that is,

$$p_{11} = p_1(p_1 + \rho p_0). \qquad (A.1)$$

Assume that

$$p_{01} = p_{10}.$$

The expression

$$p_{00} + 2p_{01} + p_{11} = 1 \qquad (A.2)$$

can be written

$$p_{00} + 2p_{0|1}\, p_1 + p_{11} = 1 \qquad (A.3)$$

but

$$p_{0|1} = 1 - p_{1|1}$$

so that (A.3) becomes

$$p_{00} + 2p_1 - p_{11} = 1$$

and substituting from (A.1), we obtain

$$p_{00} = \cancel{1 - 2p_1} + p_1^2 + \rho p_1 p_0$$

$$= p_0(p_0 + \rho p_1). \qquad \text{(A.4)}$$

Finally (A.2) may be rewritten

$$p_{01} = (1 - p_{00} - p_{11})/2$$

and substituting from (A.1) and (A.4), we have

$$p_{01} = p_0 p_1 (1 - \rho). \qquad \text{(A.5)}$$

This correlated binary distribution handles all special cases, e.g.

1. Independence($\rho = 0$):

   (A.1),(A.4), and (A.5) become

   $$p_{11} = p_1^2$$

   $$p_{00} = p_0^2,$$

   and

   $$p_{01} = p_{10} = p_0 p_1.$$

2. Cluster-specific X′s ($\rho = 1$):

   (A.1),(A.4), and (A.5) become

   $$p_{11} = p_1,$$

   $$p_{00} = p_0,$$

   $$p_{10} = p_{01} = 0.$$

3. Designed X′s (that is, $X_1 = 1$ iff $X_2 = 0$, and visa versa; hence, $\rho = -1$)

   (A.1),(A.4), and (A.5) become

   $$p_{11} = p_1(p_1 - p_0),$$

$$p_{00} = p_0(p_0 - p_1),$$

$$p_{01} = p_{10} = 2p_0p_1.$$

This is a valid probability distribution function if and only if

$$p_0 = p_1 = 1/2.$$

Appendix B

Material pertaining to Chapter 3

B.1 Dummy variables estimator and clusters with similar Y's

In this section it is shown that, for the dummy variables approach, all clusters with similar Y's, that is, clusters with all Y's equal to 1 or all Y's equal to 0, must be excluded, leading to a distribution of the dependent variable which is different from that considered for the maximum likelihood, conditional and usual estimators.

Consider the estimating equations for the dummy variables approach as defined in Section 3.3.4, and in particular, the equation (3.3.4.5) for the i'th cluster which has all Y values equal to 1. For such a cluster the equation becomes

$$n = \sum_{j=1}^{n} p_{ij}$$

where p_{ij} is the estimated probability that unit j in cluster i has a Y value of 1. This equation implies that

$$p_{ij} = 1, \; j=1,\dots,n,$$

that is,

$$\frac{\exp(\hat{\boldsymbol{\beta}}_d{}'\mathbf{x}_{ij}+\hat{\gamma}_i)}{1+\exp(\hat{\boldsymbol{\beta}}_d{}'\mathbf{x}_{ij}+\hat{\gamma}_i)} = 1, \qquad (B.1.1)$$

which implies that

$$\hat{\gamma}_i \longrightarrow \infty. \qquad (B.1.2)$$

Similarly, when all the Y values in a cluster, say cluster j, are 0, we have

$$\hat{\gamma}_j \longrightarrow -\infty. \qquad (B.1.3)$$

In either case, for the cluster under consideration the estimation of the dummy variables (and hence nuisance) parameter, γ, dominates the estimation of the regression coefficient, β. In other words, the contribution of these clusters to the log likelihood is flat as far as the parameter β is concerned.

Hence for any cluster with all Y values similar, there is no information about β and that cluster can be discarded from the estimation.

In addition any numerical minimization method in which either (B.1.2) or (B.1.3) or both occurs, the solution is considered unbounded and no numerical value is returned for any of the parameter estimates. Thus one must discard the clusters with similar values of Y from the numerical calculation or no solution to the estimating equations is possible.

The result of the removal of some clusters is to change the distribution of the dependent outcome variable. The distribution of the new variable, denoted as Z, is discussed in Appendix B.2.

B.2 Distribution of outcome variables in dummy variables approach.

The distribution of the outcome variable Z in the dummy variables approach is truncated with respect to the distribution of the outcome variable Y in the other approaches. The probability distribution function of the random vector Z may be written

$$f(\mathbf{z}) = f(\mathbf{y})c, \quad \mathbf{y} \neq \mathbf{0} \text{ or } \mathbf{1}$$
$$= 0, \text{ otherwise}$$

where

$$c = 1/(1-g_0-g_n),$$

where g_i has been defined in section 3.4.3 and

$$g_0 = \Pr(\text{no } Y\text{'a are } 1)$$
$$g_n = \Pr(n\ Y\text{'s are } 1).$$

We are most interested in the relationship in between the moments of the distributons of Z and Y, in particular the means, variances and covariances.

$$E(Z) = \Pr(Z=1)$$
$$= (f_1-g_n)c$$

where f_1 is the marginal probability of Y. This expression may be written as

$$f_1 + (f_1 g_0 - f_0 g_n)c$$

where

$$f_0 = 1 - f_1.$$

Hence

$$E(Z) = E(Y) + (f_1 g_0 - f_0 g_n)c. \qquad \text{(B.2.1)}$$

Further

$$\begin{aligned} Var(Z) &= E(Z^2) - E(Z)^2 \\ &= Pr(Z=1)[1 - Pr(Z=1)] \\ &= (f_1-g_n)c\ [1-(f_1-g_n)c] \\ &= (f_1-g_n)(f_0-g_0)\ c^2 \\ &= [f_0f_1 - (f_1g_0 + f_0g_n - g_0g_n)]c^2 \\ &= [Var(Y) - (f_1g_0 + f_0g_n - g_0g_n)]c^2. \quad (B.2.2) \end{aligned}$$

Next

$$\begin{aligned} Cov(Z_1,Z_2) &= E(Z_1Z_2) - E(Z_1)E(Z_2) \\ &= Pr(Z_1=1,Z_2=1) - [E(Z)]^2 \\ &= (f_{11} - g_n)c - (f_1 - g_n)^2c^2 \end{aligned}$$

where

$$f_{11} = Pr(Y_1=Y_2=1)$$

so that

$$\begin{aligned} cov(Z_1,Z_2) &= [(f_{11}-g_n)(1-g_0-g_n) - (f_1-g_n)(f_1-g_n)]c^2 \\ &= [f_{11}-f_1^2 - f_{11}(g_0+g_n) - g_n + g_n(g_0+g_n) \\ &\qquad + 2f_1g_n - g_n^2]c^2 \\ &= [cov(Y_1,Y_2) - f_{11}(g_0+g_n) - f_0g_n + g_n(f_1+g_0)]c^2 \\ &= [cov(Y_1,Y_2) - f_{11}(g_0+g_n) + (f_1-f_0+g_0)g_n]c^2. \end{aligned}$$

(B.2.3)

Now the relationship between $\rho_z(=corr(Z_1,Z_2))$ and ρ_y can be established from (B.2.2) and (B.2.3).

These expressions simplify somewhat for

$$f_0 = f_1 = 0.5$$

and

$$g_0 = g_n = g$$

Then we have

$$c = 1/(1-2g)$$

and (B.2.1), (B.2.2) and (B.2.3) become

$$E(Z) = E(Y)$$

$$\mathrm{Var}(Z) = [\mathrm{Var}(Y) - g + g^2]c^2$$

and

$$\mathrm{cov}(Z_1,Z_2) = [\mathrm{cov}(Y_1,Y_2) - 2\, f_{11}g + g^2]c^2.$$

B.3 Distribution of dummy variables random variable P

P is the estimated probability, under the dummy variables model, that the dependent variable has value 1, that is, for unit j in cluster i,

$$p_{ij} = \exp(\hat{\boldsymbol\beta}_d{}'\mathbf{x}_{ij}+\hat{\gamma}_i)/[1+\exp(\hat{\boldsymbol\beta}_d{}'\mathbf{x}_{ij}+\hat{\gamma}_i)].$$

Now, in section 3.3.2, it was shown that, when

$$\mathbf{b} = \mathbf{0}, \qquad (B.3.1)$$

$$k \longrightarrow \infty \Longrightarrow \hat{\boldsymbol\beta}_d \longrightarrow \mathbf{0}$$

so that p_{ij} becomes

$$p_i = \exp(\gamma_{0i})/[1+\exp(\gamma_{0i})],$$

where γ_{0i} is the limiting value of $\hat{\gamma}_i$. Now the estimating equations require that

$$m_{1i} = n\, p_i,$$

where m_{1i} is the number of units in cluster i with Y value of 1. Thus p_i can have any of (n-1) values, namely, 1,...,n-1. The values 0 and n do not occur for the reasons explained in Appendix B.1. Hence, the limiting distribution of P has only n-1 possible values of p, that is,

$$p = r/n, \ r=1,\ldots,n,$$

and the probability of each one of these values depends on the limiting distribution of R, where

$$R = \sum_{j=1}^{n} Y_j .$$

But R, under the initial assumption (B.3.1), has the probability function

$$Pr(R=r) = g_r \, c$$

where g_r was defined in section 3.4.3 and c was defined in Appendix B.2.

B.4 Distribution of an estimator under the wrong model

The asymptotic distribution of an estimator, $\hat{\boldsymbol{\theta}}$, in this case, the maximum likelihood estimator, can be obtained by expanding the estimating equations, $\mathbf{S}_k(\hat{\boldsymbol{\theta}})$, in this case, the maximum likelihood equations, in a multivariate Taylor series expansion about the correct value, $\boldsymbol{\theta}$, of the parameter (see Cox and Hinkley (1974), p.294-302). Hence

$$\mathbf{S}_k(\hat{\boldsymbol{\theta}}) = \mathbf{S}_k(\boldsymbol{\theta}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' S_k'(\boldsymbol{\theta}), \qquad \text{(B.4.1)}$$

where $S_k'(.)$ is the matrix of first derivatives of $S_k(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. We are ignoring terms of $o_p(k^{-1/2})$.

Since

$$S_k(\hat{\theta}) = 0,$$

(B.4.1) becomes

$$(\hat{\theta} - \theta) = -[S_k'(\theta)]^{-1} S_k(\theta)$$

or

$$\sqrt{k}(\hat{\theta} - \theta) = k[S_k'(\theta)]^{-1} [-S_k(\theta)/\sqrt{k}]$$

In the usual case, that is, where we have the correct model, by the multivariate central limit theorem,

$$\frac{1}{\sqrt{k}} S_k(\theta) \xrightarrow{D} MVN(0, I),$$

where I is the Fisher Information matrix. Moreover

$$-\frac{1}{k} S_k'(\theta) \xrightarrow{P} I$$

so that

$$\sqrt{k}(\hat{\theta} - \theta) \xrightarrow{D} MVN(0, I^{-1})$$

However, in the case of an incorrectly specified probability distribution,

$$\hat{\theta} \longrightarrow \theta_0$$

where θ_0 is usually not θ, the value under the correct model, so that we must examine the behaviour of

$$\sqrt{k}(\hat{\theta} - \theta_0) = k[S_k'(\theta_0)]^{-1} [-S_k(\theta_0)/\sqrt{k}]$$

under the correct model.

The following steps must be taken

1. calculate $\boldsymbol{\theta}_0$

2. evaluate $E[\mathbf{S}_k(\boldsymbol{\theta}_0)]$ and $Var[\mathbf{S}_k(\boldsymbol{\theta}_0)]$, so that, if

$$E[\mathbf{S}_k(\boldsymbol{\theta}_0)] = \mathbf{0}$$

and

$$Var[\mathbf{S}_k(\boldsymbol{\theta}_0)] = V,$$

then, by the multi-variate central limit theorem,

$$\frac{1}{\sqrt{k}} S_k(\boldsymbol{\theta}_0) \xrightarrow{D} MVN(\mathbf{0}, V),$$

(where V is not the Fisher's information matrix for either the correct or the incorrect model).

3. evaluate

$$U_k = S_k'(\boldsymbol{\theta}_0)/k$$

and obtain its limiting value U

4. Then, since,

$$\sqrt{k}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} MVN(\mathbf{0}, U^{-1}VU^{-1}),$$

we are interested in calculating all or part of

$$U^{-1}VU^{-1}.$$

B.5 Calculation of the unconditional asymptotic variance

The calculation of an unconditional variance follows this argument. If the information in a likelihood assuming condition 1 is denoted by I_1 and the information under condition 2 is I_2, and the probability of conditions 1 and 2 are denoted by p_1 and p_2, respectively, then the total information is measured by

$$I_1p_1 + I_2p_2.$$

For the dummy variables estimator, condition 1 is that the Y values are dissimilar, which occurs with probability

$$1 - g_0 - g_n$$

where g_0 and g_n have been defined in Appendix B.2. The information given in this case is assumed to be the inverse of the variance-covariance matrix given by expression (3.4.5.8). Condition 2 is that the Y values are all 1 or all 0, which occurs with probability

$$g_0 + g_n$$

but has information content 0. Hence the information in the dummy variables approach is merely

$$I_1p_1$$

and the asymptotic variance of the dummy variables estimator is given by the inverse of this information which is the matrix given by expression (3.4.5.8) divided by

$$1 - g_0 - g_n$$

For simplicity, in Appendix B.2, we have defined

$$1/[1 - g_1 - g_n)$$

as c and this term is used in section 3.4.5.

Appendix C

Material Pertaining to Chapter 4

C.1 Maximum likelihood estimators for single binary covariate

For the usual logistic model, the log likelihood function is

$$y(\alpha + \beta x) - \log d$$

where y is 0 or 1 and

$$d = 1 + \exp(\alpha + \beta x).$$

For a simple random sample of N observations from this logistic model, the log likelihood can be written

$$l = \sum_{i=1}^{N} [y_i (\alpha + \beta x_i) - \log d_i]$$

where

$$d_i = 1 + \exp(\alpha + \beta x_i), \; i = 1,\ldots,N.$$

At the maximum likelihood estimates, $\hat{\alpha}$ and $\hat{\beta}$, the first derivatives of l must be zero, that is,

$$\Sigma \; y_i = \Sigma \; \hat{p}_i \qquad (C.1.1)$$

and

$$\Sigma \; x_i y_i = \Sigma \; x_i \hat{p}_i \qquad (C.1.2)$$

where

$$\hat{p}_i = \exp(\hat{\alpha} + \hat{\beta} x_i)/[1 + \exp(\hat{\alpha} + \hat{\beta} x_i)].$$

(C.1.1) may be rewritten as

$$n_{+1} = n_{0+} f_0 + n_{1+} f_1 \qquad (C.1.3)$$

where n_{ij} is the number of observations with an x value of i and a y value of j, and

$$n_{+j} = n_{0j} + n_{1j},$$

$$n_{i+} = n_{i0} + n_{i1},$$

$$f_0 = \exp(\hat{\alpha}) / [1 + \exp(\hat{\alpha})],$$

and

$$f_1 = \exp(\hat{\alpha} + \hat{\beta})/[1 + \exp(\hat{\alpha} + \hat{\beta})].$$

Similarly (C.1.2) may be rewritten as

$$n_{11} = n_{1+} f_1. \qquad (C.1.4)$$

Subtracting (C.1.4) form (C.1.3) we get

$$n_{01} = n_{0+} f_0.$$

When f_0 is written in terms of $\hat{\alpha}$, this becomes

$$n_{01} = n_{0+} \{\exp(\hat{\alpha})/[1 + \exp(\hat{\alpha})]\}$$

which, when solved for $\hat{\alpha}$ becomes

$$\hat{\alpha} = \log(n_{01}/n_{00}). \qquad (C.1.5)$$

When we write (C.1.4) in terms of $\hat{\alpha}$ and $\hat{\beta}$, we get

$$n_{11} = n_{1+} \exp(\hat{\alpha} + \hat{\beta})/[1 + \exp(\hat{\alpha} + \hat{\beta})]$$

or

$$n_{11} = n_{10} \exp(\hat{\alpha} + \hat{\beta}).$$

Substituting for $\hat{\alpha}$ from (C.1.5) and solving for $\hat{\beta}$, we get

$$\hat{\beta} = \log(n_{11}n_{00}/n_{10}n_{01})$$

C.2 Limiting value of log odds ratio estimator

The limiting value of the log odds ratio estimator when the underlying distribution is the correlated logistic has a complex algebraic expression even for the simplest case of a single binary covariate with two units per cluster. For, as

$$k \longrightarrow \infty,$$

$$\frac{n_{00}n_{11}}{n_{01}n_{10}} \xrightarrow{p} \frac{E(R_{00})E(R_{11})}{E(R_{01})E(R_{10})}$$

$$= \frac{p_{00}[P(0,0|0,0)+P(1,0|0,0)]+p_{10}[P(0,0|1,0)+P(1,0|1,0)]}{p_{00}[P(0,1|0,0)+P(1,1|0,0)]+p_{10}[P(0,1|1,0)+P(1,1|1,0)]}$$

$$\times \frac{p_{11}[P(1,0|1,1)+P(1,1|1,1)]+p_{10}[P(1,0|1,0)+P(1,1|1,0)]}{p_{11}[P(0,0|1,1)+P(1,0|1,1)]+p_{10}[P(1,0|1,0)+P(0,0|1,0)]}$$

which may be written parametrically as

$$\exp(\beta)\,\frac{\{p_{00}d_{10}[1+\exp(\alpha_1)]+p_{10}d_{00}[1+\exp(\alpha_1+\beta)]\}}{\{p_{10}d_{11}[1+\exp(\alpha_1)]+p_{11}d_{10}[1+\exp(\alpha_1+\beta)]\}}$$

$$\times \frac{\{p_{11}d_{10}[\exp(\alpha_1)+\exp(\alpha_2+\beta)]+p_{10}d_{11}[\exp(\alpha_1)+\exp(\alpha_2)]\}}{\{p_{10}d_{00}[\exp(\alpha_1)+\exp(\alpha_2+\beta)]+p_{00}d_{10}[\exp(\alpha_1)+\exp(\alpha_2)]\}} \quad \text{(C.2.1)}$$

where

$$d_{00} = 1 + 2\exp(\alpha_1) + \exp(\alpha_2) ,$$

$$d_{01} = d_{10} = 1 + \exp(\alpha_1) + \exp(\alpha_1+\beta) + \exp(\alpha_2+\beta) ,$$

and

$$d_{11} = 1 + 2\exp(\alpha_1+\beta) + \exp(\alpha_2+\beta) .$$

This expression is equal to $\exp(\beta)$ when

1. $\beta=0$, that is, when there is no effect of covariate on the dependent variable,

or

2. $\alpha_2=2\alpha_1$, that is, when there is independence of units within the cluster.

Next we assume that the distribution of the covariates is that of a simple correlated binary (see Appendix A). We can write the second term of (C.2.1), known as the inflation factor, as

$$\frac{\{d_{10}(p_0+\rho_x p_1)[1+\exp(\alpha_1)]+d_{00}p_1(1-\rho_x)[1+\exp(\alpha_1+\beta)]\}}{\{d_{11}p_0(1-\rho_x)[1+\exp(\alpha_1)]+d_{10}(p_1+\rho_x p_1)[1+\exp(\alpha_1+\beta)]\}}$$
$$\times\frac{\{d_{10}(p_1+\rho_x p_0)[\exp(\alpha_1)+\exp(\alpha_2+\beta)]+d_{11}p_0(1-\rho_x)[\exp(\alpha_1)+\exp(\alpha_2)]\}}{\{d_{00}p_1(1-\rho_x)[\exp(\alpha_1)+\exp(\alpha_2+\beta)]+d_{10}(p_0+\rho_x p_1)[\exp(\alpha_1)+\exp(\alpha_2)]\}}$$

This expression may be evaluated to yield values of the inflation factor for various values of ρ_x, ρ_y and β. It is the simplest expression available for an inflation factor.

C.3 Limiting value of the dummy variable estimator

We are assuming only a single binary covariate and clusters of size 2. By the arguments given in section 4.3 we need only consider clusters with dissimilar x values. Let us say that for cluster i we have

$$x_1 = 1$$

and

$$x_2 = 0$$

Expression (3.3.4.5) evaluated for this cluster means that

$$1 = p_{i1} + p_{i2}$$

where p_{i1} and p_{i2} are the estimated probabilities. This expression may be written

$$1 = \frac{\exp(\hat{\beta} + \hat{\gamma}_i)}{1 + \exp(\hat{\beta} + \hat{\gamma}_i)} + \frac{\exp(\hat{\gamma}_i)}{1 + \exp(\hat{\gamma}_i)}$$

that is,

$$[1 + \exp(\hat{\beta} + \hat{\gamma}_i)][1 + \exp(\hat{\gamma}_i)]$$
$$= [\exp(\hat{\beta} + \hat{\gamma}_i)][1 + \exp(\hat{\gamma}_i)]$$
$$+ [\exp(\hat{\gamma}_i)][1 + \exp(\hat{\beta} + \hat{\gamma}_i)]$$

which becomes

$$1 = \exp(\hat{\beta} + 2\hat{\gamma}_i)$$

that is,

$$\hat{\gamma}_i = -\hat{\beta}/2 \quad (C.3.2)$$

so that

$$p_{i1} = \exp(\hat{\beta}/2) \; / \; [1 + \exp\,(\hat{\beta}/2)] \quad (C.3.3)$$
$$= p,$$

say, for all these clusters, and

$$p_{i2} = 1 - p.$$

For the other clusters, that is those, with

$$x_1 = 0$$

and

$$x_2 = 1$$

we get

$$p_{i1} = 1 - p\ .$$

and

$$R_{12} = p.$$

Hence expression (3.3.4.4) evaluated over the clusters with both dissimilar x´s and dissimilar y´s yields

$$r_{11} = r_1 p, \qquad (C.3.4)$$

where r_1 is the total number of clusters with dissimilar x´s and dissimilar y´s and r_{11} is the number of these clusters within which each unit has the same x and y value. Solving (C.3.4) for p and then solving (C.3.3) for $\hat{\beta}$, we get

$$\exp(\hat{\beta}) = (r_{11}/r_{10})^2$$

where r_{10} is the number of clusters in which each unit has a different x value from its y value, that is,

$$r_{10} = r_1 - r_{11}.$$

In a manner similar to that used for the usual logistic model, it may be shown that, conditional upon the total r_1 (= $r_{11} + r_{10}$),

$$\frac{r_{11}}{r_{10}} \xrightarrow{p} \frac{\Pr(Y_1=1,Y_2=0|X_1=1,X_2=0)}{\Pr(Y_1=0,Y_2=1|X_1=1,X_2=0)}$$

which, under Rosner´s model is

$$\frac{\exp(\alpha_1+\beta)}{\exp(\alpha_1)} = \exp(\beta).$$

Hence the dummy variables approach produces an estimator of $\exp(\beta)$ with an inflation factor equal to the value being estimated.

Appendix D

Programs and subroutines

D.1 Program BIASU

For a single binary covariate, this program computes the limiting values, bias and relative bias of the usual estimator, plus the limiting value of the estimator of the odds ratio and its inflation factor, over a specified number of values of N, the cluster size, RHOX, the intraclass correlation coefficient of the covariate, RHOY, the residual intraclass correlation of Rosner's version of the correlated logistic regression model, and EBETA, the true odds ratio being estimated.

```
      PROGRAM BIASU
C
C     CALCULATE NUMERICAL VALUE OF ASYMPTOTIC BIAS OF
C     USUAL ESTIMATOR OF BETA
C     FOR VARYING VALUES OF N(CLUSTER SIZE),
C     RHOX, RHOY/X AND BETA
C
      DOUBLE PRECISION PY,PHIY,EBET,PX,PHIX,SUM
     1,INFAC(10),OR(10),BH(10),BIAS(10),RHOX(10),RHOY(10),
     2EBETA(10),BETA(10),RBIAS(10)
     3, N00,N10,N01,N11
      DIMENSION N(10)
      COMMON /PARAM/PY,PHIY,EBET
C      CHARACTER Y
      COMMON /IO/INP,IOUT1,IOUT2,IOUT3,IOUT4,IOUT5
      DATA ITTY/1/,NSIG,MAXFN/0,0/
C
      OPEN(INP,FILE='BIAS.INP')
      OPEN(IOUT1,FILE='BIAS.OUT1')
      OPEN(IOUT2,FILE='BIAS.OUT2')
      OPEN(IOUT3,FILE='BIAS.OUT3')
      OPEN(IOUT4,FILE='BIAS.OUT4')
      OPEN(IOUT5,FILE='BIAS.OUT5')
```

```
C
C     WRITE IDENTIFYING INFORMATION TO FILE
C
      READ(INP,*) IP,IC
      WRITE(IOUT1,1990) IP
      IP = IP + 1
1990  FORMAT(1H1///72X,I3)
      IF(IC.LT.10) WRITE(IOUT1,2000)IC
      IF(IC.GE.10) WRITE(IOUT1,1999)IC
2000  FORMAT(/////46X,' TABLE 4.',I1)
1999  FORMAT(/////45X,' TABLE 4.',I2)
C
      WRITE(IOUT1,2001)
2001  FORMAT(/31X,'LIMITING VALUE OF THE USUAL ESTIMATOR ')
C
      IC = IC + 1
      WRITE(IOUT2,1990) IP
      IP = IP + 1
      IF(IC.LT.10) WRITE(IOUT2,2000)IC
      IF(IC.GE.10) WRITE(IOUT2,1999)IC
      WRITE(IOUT2,2002)
2002  FORMAT(/31X,'ASYMPTOTIC BIAS OF THE USUAL ESTIMATOR ')
C
      IC = IC + 1
      WRITE(IOUT3,1990) IP
      IP = IP + 1
      IF(IC.LT.10) WRITE(IOUT3,2000)IC
      IF(IC.GE.10) WRITE(IOUT3,1999)IC
      WRITE(IOUT3,2003)
2003  FORMAT(/33X,'RELATIVE BIAS OF THE USUAL ESTIMATOR ')
C
      IC = IC + 1
      WRITE(IOUT4,1990) IP
      IP = IP + 1
      IF(IC.LT.10) WRITE(IOUT4,2000)IC
      IF(IC.GE.10) WRITE(IOUT4,1999)IC
      WRITE(IOUT4,2004)
2004  FORMAT(/29X,'LIMITING VALUE OF ESTIMATOR OF ODDS RATIO ')
C
      IC = IC + 1
      WRITE(IOUT5,1990) IP
      IF(IC.LT.10) WRITE(IOUT5,2000)IC
      IF(IC.GE.10) WRITE(IOUT5,1999)IC
      WRITE(IOUT5,2005)
2005  FORMAT(/27X,'INFLATION FACTOR FOR ESTIMATOR OF ODDS RATIO')
```

```
C
C     READ PARAMETERS
C
      READ(INP,*) NN,(N(I),I=1,NN)
      READ(INP,*) NRX,(RHOX(I),I=1,NRX)
      READ(INP,*) NRY,(RHOY(I),I=1,NRY)
      READ(INP,*) NBETA,(EBETA(I),I=1,NBETA)
C
C     CHANGE EXP(B) TO B
C
      DO 20 I = 1,NBETA
20       BETA(I) = DLOG(EBETA(I))
C
C     LOOP FOR SAMPLE SIZES (N(I))
C
      DO 100 L = 1,NN
C
      WRITE(IOUT1,2010) N(L),(BETA(II),II=1,NBETA)
      WRITE(IOUT1,2015)(EBETA(IJ),IJ=1,NBETA)
2010  FORMAT(/40X,'CLUSTERS OF SIZE ',I2
     1 //25X,' VALUES OF BETA (FOLLOWED BY VALUES OF EXP(BETA))'
     2 //34X, 10F10.3)
2015  FORMAT(33X,10F10.1)
C
      WRITE(IOUT1,1925)
1925  FORMAT(15X,' RHOX          RHOY/X'/)
C
      WRITE(IOUT2,2010) N(L),(BETA(II),II=1,NBETA)
      WRITE(IOUT2,2015)(EBETA(IJ),IJ=1,NBETA)
      WRITE(IOUT2,1925)
C
      WRITE(IOUT3,2010) N(L),(BETA(II),II=1,NBETA)
      WRITE(IOUT3,2015)(EBETA(IJ),IJ=1,NBETA)
      WRITE(IOUT3,1925)
C
      WRITE(IOUT4,2010) N(L),(BETA(II),II=1,NBETA)
      WRITE(IOUT4,2015)(EBETA(IJ),IJ=1,NBETA)
      WRITE(IOUT4,1925)
C
      WRITE(IOUT5,2010) N(L),(BETA(II),II=1,NBETA)
      WRITE(IOUT5,2015)(EBETA(IJ),IJ=1,NBETA)
      WRITE(IOUT5,1925)
```

```
C
C       LOOP FOR RHOX
C
              DO 90 I = 1,NRX
C
      WRITE(ITTY,1500) I,NRX
1500  FORMAT(' *** COUNTER FOR NUMBER OF RHOXS AT ',I2,' OF ',I2)
      WRITE(IOUT1,2030) RHOX(I)
      WRITE(IOUT2,2030) RHOX(I)
      WRITE(IOUT3,2030) RHOX(I)
      WRITE(IOUT4,2030) RHOX(I)
      WRITE(IOUT5,2030) RHOX(I)
2030  FORMAT(15X,F5.2)
C
C     LOOP FOR RHOY
C
                    DO 90 J = 1,NRY
C     LOOP FOR BETA
C
              DO 80 K=1,NBETA
C
C     SET UP VALUES FOR EVALUATION
C
      PX = 0.5
      PHIX = RHOX(I)
      PY = 0.5
      PHIY = RHOY(J)
      EBET = EBETA(K)
C
      CALL DISTXY(N(L),PX,PHIX N00,N10,N01,N11,N1)
C
      SUM = N00 + N10 + N01 + N11
      OR(K) = N00*N11/(N10*N01)
C
      BH(K) = DLOG(OR(K))
      BIAS(K) = BH(K) - BETA(K)
      IF(K.GT.1) RBIAS(K) = BIAS(K)/BETA(K)
      INFAC(K) = OR(K)/EBET
C
80    CONTINUE
C
      WRITE(IOUT1,1600) RHOY(J),(BH(K),K=1,NBETA)
      WRITE(IOUT2,1600) RHOY(J),(BIAS(K),K=1,NBETA)
      WRITE(IOUT3,1610) RHOY(J),(RBIAS(K),K=2,NBETA)
      WRITE(IOUT4,1600) RHOY(J),(OR(K),K=1,NBETA)
      WRITE(IOUT5,1600) RHOY(J),(INFAC(K),K=1,NBETA)
C
1600  FORMAT(24X,10F10.3)
1610  FORMAT(24X,F10.3,10X,10F10.3)
C
90    CONTINUE
```

```
C
100   CONTINUE
C
      CLOSE(IOUT1)
      CLOSE(IOUT2)
      CLOSE(IOUT3)
      CLOSE(IOUT4)
      CLOSE(IOUT5)
      CLOSE(INP)
      WRITE(ITTY,2200)
2200  FORMAT(' RUN COMPLETE.')
C
      CALL EXIT
      END

      BLOCK DATA
      COMMON /IO/INP,IOUT1,IOUT2,IOUT3,IOUT4,IOUT5
      DATA INP,IOUT1,IOUT2,IOUT3,IOUT4,IOUT5/5,6,7,8,9,10/
      END
```

```
      SUBROUTINE DISTXY(N,PX,RHOX,N00,N10,N01,N11,N1)
C
C     THIS SUBROUTINE CALCULATES EXPECTED NUMBER OF (X,Y) PAIRS
C
C     WHEN MARGINAL DISTRIBUTION OF X'S IS BETA-BINOMIAL
C     AND CONDITIONAL DISTRIBUTION OF Y'S IS ROSNER'S
C
      INTEGER RX;RY
      DOUBLE PRECISION N00,N10,N01,N11,R00,R10,R01,R11,SUM,F1,F2,F3
     1,BETAB,ROSNER,CHOOSE,PX,RHOX
C
      N00 = 0.0
      N10 = 0.0
      N01 = 0.0
      N11 = 0.0
C
      DO 100 RX = 0,N
           R00 = 0.0
           R10 = 0.0
           R01 = 0.0
           R11 = 0.0
           SUM = 0.0
C
C     MARGINAL DISTRIBUTION OF X
C
           F1 = BETAB(N,PX,RHOX,RX)
C
              DO 90 RY = 0,N
C
              LOR = MAX0(RY+RX-N,0)
              LUR = MIN0(RY,RX)
C
                  DO 80 I = LOR, LUR
C
C     CONDITIONAL DISTRIBUTION OF Y
C
                  F2 = ROSNER(N,RY,I)
C
                  F3 = CHOOSE(I,RX) * CHOOSE(RY-I,N-RX)  * F2
C
C     NOW CHANGE COUNTERS
C
                  R00 = R00 + DFLOAT(N-RX-RY+I) *F3
                  R10 = R10 + DFLOAT(RX-I)*F3
                  R01 = R01 + DFLOAT(RY-I)*F3
                  R11 = R11 + DFLOAT(I)*F3
                  SUM = SUM + F3
C
80                       CONTINUE
C
90                       CONTINUE
```

```
C
C
C     ADD CURRENT VALUE
C
          N00 = N00 + F1 * R00/SUM
          N10 = N10 + F1 * R10/SUM
          N01 = N01 + F1 * R01/SUM
          N11 = N11 + F1 * R11/SUM
100       CONTINUE
C
      RETURN
      END
```

```
      DOUBLE PRECISION FUNCTION ROSNER(N,R,NM)
C
C     THIS FUNCTION CALCULATES THE VALUE OF ROSNER'S DISTRIBUTION
C     OF Y ASSUMING X'S ARE BINARY
C
      INTEGER R
      DOUBLE PRECISION P,RHO,EBETA,Q,F
      COMMON /PARAM/P,RHO,EBETA
C
      Q = 1.0 - P
      NR = N - R
C
      F = EBETA**NM
C
C     CALCULATION OF EXP(ALPHA(R))
C
      IF(R.GT.0) THEN
          DO 10 I = 0,R-1
      F = F * (P + (DFLOAT(I) - P) * RHO) / (Q + (DFLOAT(NR + I) - Q)
     1         * RHO)
10        CONTINUE
      ENDIF
C
      ROSNER = F
      RETURN
      END
```

```
      DOUBLE PRECISION FUNCTION BETAB(N,P,PHI,R)
C
C     THIS FUNCTION PROVIDES THE PROBABILITY MASS FUNCTION
C     OF A BETA-BINOMIAL DISTRIBUTION
C     GENERALIZED TO ALLOW FOR NEGATIVE CORRELATION
C
      INTEGER R
      DOUBLE PRECISION NUMA,NUMB,NUMC,DEN,P,PHI,Q,CHOOSE
C
      Q = 1.0 - P
C
C     DENOMINATOR
C
      DEN =1.0
           DO 10 I = 0,N-1
           DEN = DEN *(1.0 + DFLOAT(I-1) * PHI)
10         CONTINUE
C
C     THREE TERMS OF NUMERATOR
C
      NUMA = 1.0
      IF(R.GT.0) THEN
           DO 20 I = 0,R-1
           NUMA = NUMA * (P + (DFLOAT(I) - P) * PHI)
20    CONTINUE
           ENDIF
C
      NUMB = 1.0
           IF(R.LT.N) THEN
           DO 30 I = 0,N-R-1
           NUMB = NUMB * (Q + (DFLOAT(I) - Q) * PHI)
30         CONTINUE
           ENDIF
C
      NUMC = CHOOSE(R,N)
C
      BETAB = NUMA*NUMB*NUMC/DEN
C
      RETURN
      END
```

```
      DOUBLE PRECISION FUNCTION CHOOSE(R,N)
C
C     THIS FUNCTION EVALUATES THE COMBINATORIAL N-C-R
C     IN SUCH A WAY AS TO AVOID OVERFLOW OR UNDER FLOW (IF POSSIBLE)
C
      INTEGER R
      DOUBLE PRECISION X
C
      J = MINO(R,N-R)
      X = 0.
C
      IF(R.LE.N) THEN
C
          IF(J.GT.0) THEN
          DO 10 I = 1,J
          X = X + DLOG(DFLOAT(N-I+1)/DFLOAT(J-I+1))
10        CONTINUE
          ENDIF
C
      ELSE
      WRITE(1,100) R,N
100   FORMAT(' *** ERROR IN SUBROUTINE CHOOSE ***'
     1/     ' R VALUE OF ',I3,' GREATER THAN N VALUE OF ',I3)
      ENDIF
C
      CHOOSE = DEXP(X)
      RETURN
      END

      DOUBLE PRECISION FUNCTION DFLOAT(I)
      DFLOAT = DBLE(FLOAT(I))
      RETURN
      END
```

has mean 0.5. The distributions used for Z and X were those used in the previous section, that is, Rosner's correlated logistic for Z (truncated, of course, not to allow all dependent variables of the same value within a cluster) and the beta-binomial for x.

Table 4.30 gives the value of β_0, the limiting value of $\hat{\beta}_d$, the bias and relative bias of β_0, the limiting value of the odds ratio estimator based on the dummy variables estimator, $\exp(\beta_0)$, and the inflation factor for $\exp(\beta_0)$, for various values of β, $\rho_{y|x}$ and ρ_x and for cluster sizes

$$n = 2,3,4\ 10,25,50.$$

From the table it can be seen that

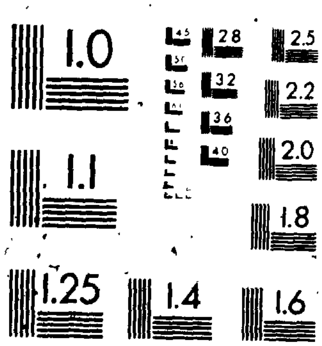
1. there is no effect of the correlation structure, namely, ρ_x and $\rho_{y|x}$, on the bias as the cluster size changes.

2. the bias and relative bias increase with the absolute value of β, although the relative bias increases only a small amount.

3. the bias and relative bias decrease as n increases, but only at a slow rate. For example, the relative bias is slightly under 5 percent for clusters of size 25, but is still greater than 2 percent for size 50.

4 OF/DE 4

MICROCOPY RESOLUTION TEST CHART
NBS 1010a
ANSI and ISO TEST CHART No 2



```
      FUNCTION EXPECT(F)
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
C
      SUM = 0.0
      DO 10 I=0,1
         DO 10 J=0,1
         SUM = SUM + F(I,J)*P(I,J)
10    CONTINUE
C
      EXPECT = SUM
      RETURN
      END

      FUNCTION EVALR()
      DIMENSION A(6)
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
      COMMON /IO/IIN,IOUT,ITTY
      COMMON /ALPHAS/EAL1,EAL2,BE
C
      EXTERNAL A1,A2,A3
C
      A(1) = EXPECT(A1)
      A(2) = EXPECT(A2)
      A(3) = EXPECT(A3)
C
      A(4) = 2.0* (F10(1,0)*F00(0,1)*P(1,0) + F10(1,1)*F00(1,1)*P(1,1)
     1  -F01(1,0)*F11(1,0)*P(1,0) - F01(1,1)*F11(1,1)*P(1,1))
      A(5) = 2.0*(F0P(1,0)*F11(1,0)*P(1,0) + F0P(1,1)*F11(1,1)*P(1,1))
      A(6) = 2.0*(F1P(1,0)*F0P(1,0)*P(1,0) + F1P(1,1)*F0P(1,1)*P(1,1)
     1 + (F11(1,1) - F1P(1,1)**2)*P(1,1))
C
      D = A(1)*A(3) - A(2)*A(2)
      WRITE(IOUT,100) A,D
100   FORMAT(' ELEMENTS OF MATRIX '/6G12.5/' DETERMINANT OF UPPER SUB'
     1/ ,G12.5)
C
      A(6) = A(6) - (A(3)*A(4)**2 - 2.0*A(2)*A(4)*A(5) + A(1)*A(5)**2)
     1 /D
      EVALR = 1.0/A(6)
      RETURN
      END
```

```
C
      WRITE(IOUT1,2001)
 2001 FORMAT(/15X,'LIMITING VALUE OF DUMMY VARIABLES ESTIMATOR ')
      WRITE(IOUT2,2002)
 2002 FORMAT(/15X,'ASYMPTOTIC BIAS OF DUMMY VARIABLES ESTIMATOR ')
      WRITE(IOUT3,2003)
 2003 FORMAT(/15X,'RELATIVE BIAS OF DUMMY VARIABLES ESTIMATOR ')
      WRITE(IOUT4,2004)
 2004 FORMAT(/20X,'LIMITING VALUE OF ESTIMATOR OF ODDS RATIO ')
      WRITE(IOUT5,2005)
 2005 FORMAT(/20X,'INFLATION FACTOR FOR ESTIMATOR OF ODDS RATIO')
C
C     READ PARAMETERS
C
      READ(INP,*) NN,(N(I),I=1,NN)
      WRITE(ITTY,990) NN,(N(I),I=1,NN)
 990  FORMAT( I5,' VALUES OF N, NAMELY '/10I5)
C
      READ(INP,*) NBETA,(EBETA(I),I=1,NBETA)
      WRITE(ITTY,1020) NBETA,(EBETA(I),I=1,NBETA)
 1020 FORMAT( I5,' VALUES OF EXP(BETA), NAMELY '/10F10.3)
C
C     CHANGE EXP(B) TO B
C
      DO 20 I = 1,NBETA
 20      BETA(I) = DLOG(EBETA(I))
C
      WRITE(IOUT1,2010) (BETA(II),II=1,NBETA)
      WRITE(IOUT1,2015)(EBETA(IJ),IJ=1,NBETA)
 2010 FORMAT(
     1 //20X,' VALUES OF BETA (FOLLOWED BY VALUES OF EXP(BETA))'
     2 //22X, 10F10.3)
 2015 FORMAT(21X,10F10.1)
C
      WRITE(IOUT1,1925)
 1925 FORMAT('     CLUSTER SIZE')
      WRITE(IOUT2,1925)
      WRITE(IOUT3,1925)
      WRITE(IOUT4,1925)
      WRITE(IOUT5,1925)
```

```
C
C     LOOP FOR SAMPLE SIZES (N(I))
C
      DO 100 L = 1,NN
C
C     LOOP FOR BETA
C
              DO 80 K=1,NBETA
C
C     SET UP VALUES FOR EVALUATION
C
      PX = 0.5
      RHOX = 0.0
      PY = 0.5
      RHOY = 0.0
      NUM = N(L)
      EBET = EBETA(K)
C
      CALL ESTIM(OR(K))
C
      WRITE(ITTY,1130) OR(K)
1130   FORMAT(' LIMITING VALUE OF ODDS RATIO ',F10.3)
C
      BH(K) = DLOG(OR(K))
      BIAS(K) = BH(K) - BETA(K)
      IF(BETA(K).GT.0) REL(K) = BIAS(K)/BETA(K)
      INFAC(K) = OR(K)/EBET
C
80    CONTINUE
C
      WRITE(IOUT1,1600) N(L),(BH(K),K=1,NBETA)
      WRITE(IOUT2,1600) N(L),(BIAS(K),K=1,NBETA)
      WRITE(IOUT3,1610) N(L),(REL(K),K=2,NBETA)
      WRITE(IOUT4,1600) N(L),(OR(K),K=1,NBETA)
      WRITE(IOUT5,1600) N(L),(INFAC(K),K=1,NBETA)
1600  FORMAT(5X,I5,12X,10F10.3)
1610  FORMAT(5X,I5,22X,9F10.3)
C
100   CONTINUE
```

```
C
      CLOSE(IOUT1)
      CLOSE(IOUT2)
      CLOSE(IOUT3)
      CLOSE(IOUT4)
      CLOSE(IOUT5)
      CLOSE(INP)
      WRITE(ITTY,2200)
2200  FORMAT(' RUN COMPLETE.')
C
      CALL EXIT
      END

      BLOCK DATA
      COMMON /IO/INP,IOUT1,IOUT2,IOUT3,IOUT4,IOUT5
      DATA INP,IOUT1,IOUT2,IOUT3,IOUT4,IOUT5/5,6,7,8,9,10/
      END

      SUBROUTINE ESTIM(BH)
      DOUBLE PRECISION BH,FUN,SMALL,ST,B
      EXTERNAL FUN
C
      SMALL = 1.0D-7
      B = 1.05
      ST = 0.005
C
      CALL SINMIN(B,ST,FUN,SMALL)
C
      BH = B
      RETURN
      END
```

```
      SUBROUTINE SINMIN(B,ST,FUN,SMALL)
C
C     THIS SUBROUTINE MINIMIZES A FUNCTION OF ONE VARIABLE
C     USING A SUCCESS-FAILURE LINEAR SEARCH WITH PARABOLIC INVERSE
C     INTERPOLATION
C     AS RECOMMENDED BY NASH(1979, P.124-30).
C
C     B IS STARTING VALUE (INPUT)
C           AND VALUE AT WHICH MINIMUM OCCURS (OUTPUT)
C     ST IS INITIAL STEP SIZE (INPUT)
C           AND FINAL STEP SIZE (OUTPUT)
C     FUN IS FUNCTION TO BE MAXIMIZED
C     SMALL IS DESIRED ACCURACY OF B
C
C     THIS IS THE DOUBLE PRECISION VERSION
C
      DOUBLE PRECISION A1,A2,EPS,BIG,B,ST,FUN,SMALL,P,S1,S0,X1,BMIN
     1 ,X2,X0
      COMMON /FUNERR/IER
      LOGICAL IER
C
      DATA A1,A2,EPS,BIG/1.5,-.25,1.0D-7,1.0D36/
C
C     IER = .FALSE.
      IF( SMALL .LT. EPS ) SMALL = EPS
C
C     INITIAL VALUE
      IFN = 1
      P = FUN(B)
20    S1 = P
      S0 = -BIG
      X1 = 0
      BMIN = B
C
30    X2 = X1 + ST
      B = BMIN + X2
C
C     TEST FOR CONVERGENCE
C     IF( (ABS(B) + SMALL) .EQ. (ABS(BMIN) + ABS(X1) + SMALL) )GO TO
250
      IF( DABS( B - (BMIN + X1) ) .LT.SMALL) GO TO 250
C
      IFN = IFN + 1
      P = FUN(B)
      IF(IER) GO TO 90
C
C     IF SUCCESS JUMP
      IF(P.LT.S1) GO TO 100
C
C     IF THIRD POINT FOUND GO TO INVERSE PARABOLIC INTERPOLATION
C
```

```
      IF(SO.GE.SI) GO TO 110
C
C     FAILURE
C
      S0 = P
      X0 = X2
C
90    ST = ST *A2
      GO TO 30
C
C     ACTION ON SUCCESS
C
100   X0 = X1
      S0 = S1
      X1 = X2
      S1 = P
      ST = ST * A1
      GO TO 30
C
C     CALCULATION OF INVERSE POLLYNOMIAL
C
110   X0 = X0 -X1
      S0 = (S0 - S1) * ST
      P = (P - ST) * X0
      IF(P.EQ.S0) GO TO 180
      ST = 0.5 * (P * X0 - S0 * ST)/(P - S0)
      X2 = X1 + ST
      B = BMIN + X2
C
      IF(B.EQ.(BMIN + X1 ) ) GO TO 200
      IFN = IFN + 1
      P = FUN(B)
      IF(IER) GO TO 190
C
      IF( P .LT. S1 ) GO TO 190
C
180      B = BMIN + X1
      P = S1
      GO TO 200
C
190   X1 = X2
200   ST = A2 * ST
      GO TO 20
C
250         RETURN
      END
```

```
      DOUBLE PRECISION FUNCTION FUN(BH)
C
C     THIS FUNCTION CALCULATES THE FIRST MAXIMUM LIKELIHOOD
C     EQUATION SQUARED, SUCH THAT,HOPEFULLY,
C     THE SUBROUTINE SINMIN WILL MINIMIZE IT AND PRODUCE
C     THE DUMMY VARIABLES ESTIMATOR OF BETA
C
C     WHEN MARGINAL DISTRIBUTION OF X'S IS BETA-BINOMIAL
C     AND CONDITIONAL DISTRIBUTION OF Y'S IS ROSNER'S
C
      INTEGER RX,RY
      COMMON/PARAM2/N,PX,RHOX
      DOUBLE PRECISION N10,N11,R10,R11,SUM0,F1,F2,F3,R1
     1,BETAB,ROSNER,CHOOSE,PX,RHOX,BH,G,SUM1,SUM2
C
      N10 = 0.0
      N11 = 0.0
      SUM2 = 0.0
C
      DO 100 RX = 1,N-1
         R10 = 0.0
         R11 = 0.0
         SUM1 = 0.0
         SUM0 = 0.0
C
C     MARGINAL DISTRIBUTION OF X
C
         F1 = BETAB(N,PX,RHOX,RX)
C
            DO 90 RY = 1,N-1
C
            LOR = MAX0(RY+RX-N,0)
            LUR = MIN0(RY,RX)
            R1 = 0.0
C
                DO 80 I = LOR, LUR
C
C     CONDITIONAL DISTRIBUTION OF Y
C
                F2 = ROSNER(N,RY,I)
C
                F3 = CHOOSE(I,RX) * CHOOSE(RY-I,N-RX)   * F2
C
C     NOW CHANGE COUNTERS
C
                R10 = R10 + DFLOAT(RX-I)*F3
                R11 = R11 + DFLOAT(I)*F3
                R1 = R1 + DFLOAT(RX)*F3
                SUM0 = SUM0 + F3
C
80                      CONTINUE
```

```
C
      SUM1 = SUM1 + R1 * G(BH,N,RX,RY)
C
90                        CONTINUE
C
C     ADD CURRENT VALUE
C
        N10 = N10 + F1 * R10/SUMO
        N11 = N11 + F1 * R11/SUMO
       SUM2 = SUM2 + F1 * SUM1/SUMO
100     CONTINUE
C
      FUN = (N11 - SUM2)**2
C
      RETURN
      END

      DOUBLE PRECISION FUNCTION G(BH,N,IX,IY)
      DOUBLE PRECISION BH,A,B,C,E,RX,RY,RN
C
      RX = DFLOAT(IX)
      RY = DFLOAT(IY)
      RN = DFLOAT(N)
C
      A = (RN - RY) * BH
      B = (RN - RY - RX) + (RX - RY) * BH
      C = -RY
C
      E = (-B + DSQRT(B*B - 4.0*A*C))/ (2.0*A)
      G = (E*BH) / (1 + E*BH)
      RETURN
      END
```

D.3 Subroutine RELEFF

For a single binary covariate with 2 units per cluster, this subroutine calculates the relative efficiency of the conditional estimator compared to the unconditional maximum likelihood estimator for any value of RHOY, the residual intraclass correlation coefficient of Rosner's version of the correlated logistic model, RHOX, the intraclass coefficient of the covariate, and BETA, the true value of the regression coefficient being estimated.

```
      FUNCTION RELEFF(RHOY,RHOX,BETA)
C
C     CALCULATE RELATIVE EFFICIENCY OF
C     CONDITIONAL ESTIMATOR
C     FOR ANY VALUE OF BETA, RHOX AND RHOY
C
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
      COMMON /ALPHAS/EAL1,EAL2,BE
       COMMON/VALUES/RHO,P0,P1
      COMMON /IO/IIN,IOUT,ITTY
      EXTERNAL EVALR,EVALC
C
C     INITIALIZATION
C
      BE = BETA
      RHO = RHOX
C
      IF(BE.EQ.0.0.AND.RHO.EQ.-1) THEN
           RELEFF = 1.0
      ELSE
C
      P1 = 0.5
      CALL XDIST
C
      EAL1 = (1.0-RHOY)/(1.0+RHOY)
      EAL2 = 1.0
      WRITE(IOUT,1860) EAL1,EAL2
1860  FORMAT(' VALUE OF EAL1 AND EAL2 '/2G12.4)
C
      WRITE(IOUT,2000) BETA
```

```
2000  FORMAT(' VALUE OF COEFFICIENT IS ',G12.4)
C
      AVARR = EVALR()
C
      AVARC = EVALC()
C
      RELEFF = AVARR/AVARC
      WRITE(IOUT,2010) AVARR,AVARC,RELEFF
2010  FORMAT(' VARIANCE OF MLE  VARIANCE OF CONDITONAL '
     1/  2(5X,G12.5)
     2/' WITH RELATIVE EFFICIENCY ',G12.5)
C
1500  FORMAT(A1)
      ENDIF
      RETURN
      END

      SUBROUTINE XDIST
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
      COMMON /IO/IIN,IOUT,ITTY
       COMMON/VALUES/RHOX,P0,P1
      P0 = 1.0 - P1
      P(0,0) = P0*(P0+RHOX*P1)
      P(0,1) = P0*P1*(1-RHOX)
      P(1,0) = P0*P1*(1-RHOX)
      P(1,1) = P1*(P1+RHOX*P0)
C
      WRITE(IOUT,100) P(0,0),P(0,1),P(1,0),P(1,1)
100   FORMAT(' DISTRIBUTION OF XS '/4F10.3)
C
      RETURN
      END
```

```
      FUNCTION EXPECT(F)
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
C
      SUM = 0.0
      DO 10 I=0,1
           DO 10 J=0,1
           SUM = SUM + F(I,J)*P(I,J)
10    CONTINUE
C
      EXPECT = SUM
      RETURN
      END

      FUNCTION EVALR()
      DIMENSION A(6)
      DIMENSION P(0:1,0:1)
      COMMON/XDIS/P
      COMMON /IO/IIN,IOUT,ITTY
      COMMON /ALPHAS/EAL1,EAL2,BE
C
      EXTERNAL A1,A2,A3
C
      A(1) = EXPECT(A1)
      A(2) = EXPECT(A2)
      A(3) = EXPECT(A3)
C
      A(4) = 2.0* (F10(1,0)*F00(0,1)*P(1,0) + F10(1,1)*F00(1,1)*P(1,1)
     1  -F01(1,0)*F11(1,0)*P(1,0) - F01(1,1)*F11(1,1)*P(1,1))
      A(5) = 2.0*(F0P(1,0)*F11(1,0)*P(1,0) + F0P(1,1)*F11(1,1)*P(1,1))
      A(6) = 2.0*(F1P(1,0)*F0P(1,0)*P(1,0) + F1P(1,1)*F0P(1,1)*P(1,1)
     1 + (F11(1,1) - F1P(1,1)**2)*P(1,1))
C
      D = A(1)*A(3) - A(2)*A(2)
      WRITE(IOUT,100) A,D
100   FORMAT(' ELEMENTS OF MATRIX '/6G12.5/' DETERMINANT OF UPPER SUB'
     1/ ,G12.5)
C
      A(6) = A(6) - (A(3)*A(4)**2 - 2.0*A(2)*A(4)*A(5) + A(1)*A(5)**2)
     1 /D
      EVALR = 1.0/A(6)
      RETURN
      END
```

```
      FUNCTION  EVALC()
      COMMON /IO/IIN,IOUT,ITTY
      COMMON/VALUES/RHOX,P0,P1
      COMMON /ALPHAS/EAL1,EAL2,BE
      EXTERNAL F01
C
      EB = EXP(BE)
      P0 = 1-P1
      V = (2*EB*P1*P0*(1-RHOX))/(1+EB)**2
      PR = 2.0*EXPECT(F01)
      WRITE(IOUT,100) V,PR
100   FORMAT(' CONDITIONAL VARIANCE AND PROBABILITY '/2(5X,G12.5))
      EVALC = 1.0/(V*PR)
      RETURN
      END

      FUNCTION F00(I,J)
      COMMON /ALPHAS/EAL1,EAL2,BE
      DEN(I,J) = 1+EAL1*EXP(BE*I) + EAL1*EXP(BE*J)
     1 + EAL2*EXP(BE*I+BE*J)
      F00 = 1.0/DEN(I,J)
      RETURN
      END

      FUNCTION F10(I,J)
      COMMON /ALPHAS/EAL1,EAL2,BE
      DEN(I,J) = 1+EAL1*EXP(BE*I) + EAL1*EXP(BE*J)
     1 + EAL2*EXP(BE*I+BE*J)
      F10 = EAL1*EXP(BE*I)/DEN(I,J)
      RETURN
      END

      FUNCTION F01(I,J)
      COMMON /ALPHAS/EAL1,EAL2,BE
      DEN(I,J) = 1+EAL1*EXP(BE*I) + EAL1*EXP(BE*J)
     1 + EAL2*EXP(BE*I+BE*J)
      F01 = EAL1*EXP(BE*J)/DEN(I,J)
      RETURN
      END

      FUNCTION F11(I,J)
      COMMON /ALPHAS/EAL1,EAL2,BE
      DEN(I,J) = 1+EAL1*EXP(BE*I) + EAL1*EXP(BE*J)
     1 + EAL2*EXP(BE*I+BE*J)
      F11 = EAL2*EXP(BE*I+BE*J)/DEN(I,J)
      RETURN
      END
```

```
FUNCTION FOP(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
FOP = F00(I,J)+F01(I,J)
RETURN
END

FUNCTION F1P(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
F1P = F10(I,J) + F11(I,J)
RETURN
END

FUNCTION FP0(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
FP0 = F00(I,J) + F10(I,J)
RETURN
END

FUNCTION FP1(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
FP1 = F01(I,J) + F11(I,J)
RETURN
END

FUNCTION A1(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
A1 = (F10(I,J)+F01(I,J))*(1.0-F10(I,J)-F01(I,J))
RETURN
END

FUNCTION A2(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
A2 = - (F10(I,J) + F01(I,J))*F11(I,J)
RETURN
END

FUNCTION A3(I,J)
COMMON /ALPHAS/EAL1,EAL2,BE
A3 = F11(I,J) * (1.0-F11(I,J))
RETURN
END
```

D.4 Subroutine XMISD

For a single binary covariate with no effect of the covariate (beta equal to 0), this subroutine calculates the misspecification factor of the dummy variables estimator for any value of RHOY, the residual intraclass correlation coefficient of Rosner's version of the correlated logistic model, RHOX, the intraclass coefficient of the covariate, and N, the cluster size.

```
      FUNCTION XMISD(RHOY,RHOX,N)
C
      IF(RHOY.EQ.1) THEN
      XMISD = 0.0
C
      ELSE
           S2Y = 0.25
           F11 = 0.25*(1+RHOY)
           G = PROB(RHOY,N)
           C = 1/(1-2*G)
           S2Z = (S2Y - G + G**2)*C**2
           COVZ = (S2Y*RHOY - 2*F11*G + G**2)*C**2
           RHOZ = COVZ/S2Z
           XNM = FLOAT(N-1)
           DEN =XNM*(1-RHOX)
           XN = FLOAT(N)
           XMISD = XN*C*(1+(N-1)*RHOZ*RHOX)/DEN
      ENDIF
      RETURN
      END

      FUNCTION PROB(RHO,N)
      P = 0.5
      DO 10 I = 2,N
10    P = P* 0.5 *(1+(2*I-3)*RHO)/(1+(I-2)*RHO)
      PROB = P
      RETURN
      END
```

References

Altham, P.M.E. (1976). Discrete variable analysis for individuals grouped into families. Biometrika 63, 163-9.

Altham, P.M.E. (1978). Two generalizations of the binomial distribution. Appl. Statist. 27, 162-7.

Andersen, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. Journ Roy. Statist. Soc. B 32, 283-301.

Andersen, E.B. (1973). Conditonal Inference and Models for Measuring. Mentalhygiejnisk Forlag. Copenhagen.

Anderson, J.A. (1972). Separate sample logistic discrimination. Biometrika, 43, 19-35.

Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W., and Weisberg, H.I. (1980). Statistical Methods for Comparative Studies. Wiley, New York.

Anderson, T.W. (1958) An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Anscombe, F. (1956). On estimating binomial response relations. Biometrika 43, 461-4.

Ashford, J.R, and Sowden, R.R. (1970). Multivariate probit analysis. Biometrics 26, 535-46.

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.), 158-76. Stanford University Press, Stanford.

Baker, R., and Nelder, J.A. (1978). The GLIM SYstem Release 3: Generaised Linear Interactive Modelling. Numerical Algorithms Group, Oxford.

Berkson, J. (1944). Application of the logistic function to bio-assay. Journ. Amer. Statist. Assoc. 39, 357-65.

Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bioassay with quantal responses, based on the logistic function. Journ. Amer. Statist. Assoc. 48, 565-99.

Berkson, J. (1955). Maximum likelihood and minimum χ^2 estimates of the logistic function. Journ. Amer. Statist. Assoc. 50, 130-62.

Berkson, J. (1968). Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of 'no interaction'. Biometrics 24, 75-95.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. Biometrics 19, 220-33.

Bliss, C. (1934). The method of probits. Science 79, 38-9.

Bloomfield, P., and Watson, G. (1975). The inefficiency of least squares. Biometrika 62, 121-8.

Bock, R.D. (1970). Estimating multinomial response relations. In Essays in Probability and Statistics. Univ. North Carolina Press, Chapel Hill.

Bradley, R.A., and Gart, J.J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. Biometrika 49, 205-14.

Breslow, N.E., Day, N.E., Halvorsen, K.T., Prentice, R.L., and Sabai, C. (1978). Estimation of multiple relative risk functions in matched case-control studies. Amer. Journ. Epid. 108, 299-307.

Breslow, N.E., and Day, N.E. (1980). Statistical Methods in Cancer Research. IARC: Lyon.

Breslow, N.E., and Powers, W. (1978). Are there two logistic regressions for retrospective studies. Biometrics 34, 100-5.

Brier, S.S. (1980). Analysis of contingency tables under cluster sampling. Biometrika 67, 591-6.

Brillinger, D.R. (1975). Statistical inference for stationary point processes. In Stochastic Processes and Related Topics (M.L. Puri ed.), 1-96. Academic Press, New York.

Campbell, C. (1979). Properties of ordinary and weighted least squares estimates in two stage samples. Proc. Soc. Statist. Section, Amer. Statist. Assoc., 800-5.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. Review Econ. Studies 17, 225-38.

Chanda, K.C. (1954). A note on the consistency and maxima of the roots of likelihood equations. Biometrika 41, 56-61.

Chatfield, C., and Goodhardt, G.J. (1970). The beta-binomial model for consumer purchasing behaviour. Appl. Statist. 19, 240-50.

Christensen, R. (1984). A note onordinary least squares methods for two-stage sampling. Journ. Amer. Statist. Assoc. 79, 720-3.

Cochran, W.G. (1943). Analysis of variance for percentage based on unequal numbers. Journ. Amer. Statist. Assoc. 38, 287-301.

Cohen, J.E. (1976). The distribution of the chi-squared statistic under clustered sampling from contingency tables. Statist. Assoc. 71, 665-70.

Cox, D.R. (1966). Some procedures associated with the logistic qualitative response curve. In Research Papers in Statistics: Festschrift for J. Neyman. (F.N.David ed.), p.55-71. Wiley, New York.

Cox, D.R. (1970). The Analysis of Binary Data. Methuen, London.

Cox, D.R. (1972). The analysis of multivariate binary data. Appl. Statist. 21, 113-20.

Cox, D.R., and Hinkley, D.V. (1974). Mathematical Statistics. Chapman and Hall, London.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

Cramer, H. (1947). Problems in probability theory. Ann. Math. Statist. 18, 165-93.

Crowder, M.J. (1978). Beta-binomial ANOVA for proportions. Appl. Statist. 25, 34-7.

Crowder, M.J. (1979). Inference about the intraclass correlation coefficient in the Beta-binomial ANOVA for proportions, Journ. Roy. Statist. Soc. B 41, 230-4.

Day, N.E., and Kerridge, D.F. (1967). A general maximum likelihood discriminant, Biometrics 23, 313-23.

Dempster, A.P. (1979). An overview of multivariate data analysis. Journ. Mult. Anal. 1, 316-46.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete observations (with discussion). Journ Roy. Statist. Soc. B 39, 1-38.

Donald, A. (1984). The analysis of clustered data in sets of 2×2 contingency tables. Ph.D. thesis, University of Western Ontario.

Donner, A.P. (1984). Linear regression analysis with repeated measurements. Journ. Chron. Dis. 37, 441-8.

Donner, A., and Koval, J. (1981). A multivariate analysis of family data. Amer. Journ. Epid. 114, 149-54.

Draper, N.R., and Smith, H. (1981) Applied Regression Analysis 2nd edition. Wiley, New York.

Dyke, G.V., and Patterson, H.D. (1952). Analysis of factorial arrangements when the data are proportions. Biometrics 8, 1-12.

Fabsitz, R. Feinleib, M., and Hubert, H. (1985). Regression analysis with correlated errors: an example from the NHLBI twin study. Journ. Chron. Dis. 38, 165-70.

Farewell, V.T. (1979). Some results on the estimation of logistic models based on retrospective data, Biometrika 66, 27-32.

Feller, W. (1968). An Introduction to Probability Theory and its Applications, 3rd edition. Wiley, New York.

Fisher, R. (1956). On a test of significance in Pearson°s Biometrika tables (no. 11). Journ Roy. Statist. Soc. B 18, 56-60.

Finney, D.J. (1971). Probit Analysis, 3rd edition. Cambridge University Press, Cambridge.

Fuller, W.A. (1975). Regression analysis for sample surveys. Sankhya 37, 117-32.

Gaddum, J. (1933). Reports in biological standards III. Methods of biological assay depending on a quantal response. Spec. Rep. Ser. Med. Coun., Lond., No. 183.

Goodman, L.A. (1963a). On Plackett's test for contingency table interactions. Journ Roy. Statist. Soc. B 25, 179-88.

Goodman, L.A. (1963b). On methods for comparing contingency tables. Journ Roy. Statist. Soc. A 126, 94-108.

Gourieroux, C., and Montfort, A. (1981). Asymptotic properties of the maximum likelihood estimatoer in dichotomous logit models. Journ. Econometrics 17, 83-97.

Graybill, F.A. (1976). Theory and Applications of the Linear Model. Duxbury, North Scituate.

Griffiths, D.A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29, 637-48.

Grizzle, J.E. (1961). A new method for testing hypotheses and estimating the parameters for the logistic model. Biometrics 17, 372-85.

Grizzle, J.E. (1971). Multivariate logit analysis. Biometrics 27, 1057-62.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics 25, 489-504.

Halperin, M. (1951). Normal regression theory in the presence of intra-class correlation. Ann. Math. Statist. 22, 573-80.

Halperin, 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. Journ. Chron. Dis. 24, 125-58.

Haseman, J.K., and Kupper, L.L. (1979) Analysis of dichotomous response data from certain toxicological experiments. Biometrics 35, 281-93.

Hauck, W. (1981). Logit analysis, preliminary draft. To be published by Wiley, New York.

Hodges, J.L. (1958). Fitting the logistic by maximum likelihood. Biometrics 14, 453-6.

Holt, D., Smith, T.M.F., and Winter, P.D. (1980). Regression analysis of data from complex surveys. Journ Roy. Statist. Soc. A 143, 474-87.

Ishii, G., and Hayakawa, R. (1960). On the compound binomial distribution. Ann. Inst. Statist. Math. 12, 69-80.

Johnson, N.L., and Kotz, S. (1969). Discrete Distributions. Houghton Mifflin, Boston.

Kempthorne, O., and Tandon, O.B. (1953). The estimation of heritability by regression of offspring on parent. Biometrics 9, 90-100.

Kleinbaum, D.G., and Kupper, L.L. (1978). Applied Regression Analysis and Other Multivariable Methods. Duxbury, North Scituate.

Kendall, M.G., and Stuart, A. (1973). The Advanced Theory of Statistics, Vol. II, 3rd edition. Griffin, London.

Kish, L., and Frankel, M.R. (1974). Inference from complex surveys. Journ Roy. Statist. Soc. B 36, 1-22.

Kleinmam, J.C. (1973). Proportions with extraneous variance: single and independent samples. Journ. Amer. Statist. Assoc.

Kleinman, J.C. (1975). Proportions with extraneous variance: two dependent samples. Biometrics 31, 737-43.

Knott, M. (1975). On the minimum efficiency of least squares. Biometrika 62, 129-32.

Konijn, H.S. (1962). Regression analysis in sample surveys. Journ. Amer. Statist. Assoc. 590-606.

Korn, E.L., and Whittemore, A.S. (1979). Methods for analysing panel studies of acute health effects of air pollution. Biometrics 35, 795-802.

Kulperger, R. (1979). Parametric estimation for a simple branching diffusion process. Journ. Mult. Anal. 9, 101-5.

Kupper, L.L., and Haseman, J.K. (1978). The use of a correlated binomial model for tha analysis of certain toxicological experiments. Biometrics 34, 69-76.

Laird, N., and Ware, T. (1981). Random effects models for serial observartions with dichotomous response, Annual meeting Amer. Statist. Assoc. Detroit, Mich.

Liddell, F.D.K., McDonald J.C., and Thomas, D.C. (1977). Methods of cohort analysis: appraisal by application to asbestos mining (with discussion). Journ Roy. Statist. Soc. A 140, 469-91.

Lubin, J.H. (1981). An empirical evaluation of the use of conditional and unconditional likelihoods in case-control data. Biometrika 49, 250-52.

Manly, B.F.J. (1978). Regression models for proportions with extraneous variance. Biometrie-Praximetrie 17, 1-18.

Mantel, N. (1966). Models for conplex contingency tables and polychotomous response response curves. Biometrics 22, 83-110.

McCullagh, P., and Nelder, J.A. (1983). Generalized Linear Models. Chapman and Hall, London.

McElroy, F.W. (1967). A necessary and sufficient condition that the ordinary least squares estimators be best linean unbiased. Journ. Amer. Statist. Assoc. 62, 1302-4.

Mossiman, J.E. (1962). On a compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. Biometrika 49, 65-82.

Nathan, G., and Holt, D. (1980). The effect of survey design on regression analysis. Journ Roy. Statist. Soc. B 42, 377-86.

Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. Journ Roy. Statist. Soc. A 135, 370-84.

Nerlove, M., and Press, S.J. (1973). Univariate and Multivariate Log-linear and Logistic Models. Rand, Santa Monica.

Neyman, J., and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica 16, 1-32.

Ochi, Y. and Prentice, R.L. (1984). Likelihood inference in a correlated probit regression model. Biometrika 71, 531-43.

Paul, S.R. (1984). A three-parameter generalization of the binomial distribution. Windsor Mathematics Report, WMR 84-11.

Pierce, D.A. and Sands, B.R. (1975). Extra-bernoulli variation in binary data. Oregon State University Statistics Department Tehnical Report, No. 46.

Prentice, R.L. (1976). Use of the logistic model in retrospectvive studies. Biometrics 32, 599-606.

Prentice, R.L., and Breslow, N.E. (1978). Retrospective studies and failure time models. Biometrika 65, 153-8.

Prentice, R.L., and Pyke, R. (1979). Logistic disease incidence models and case-control studies. Biometrika 66, 403-11.

Press, S.J. (1972). Applied Multivariate Analysis. Holt, Rinehart, and Winston, New York.

Rao, C.R. (1973). Linear Statistical Inference and its Applications. Wiley, New York.

Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence of two-way tables, Journ. Amer. Statist. Assoc. 75, 221-30.

Rao, J.N.K., and Scott, A.J. (1983). On chisquare tests for multiway contingency tables with cell proportions estimated from survey data. Carleton Math. Series No. 199, University of Carleton, Ottawa.

Rosner, B. (1983). Multivariate methods in opthamology, Annual meeting Amer. Statist. Assoc., Toronto, Ontario.

Rosner, B. (1984). Multivariate methods in opthamology with applications to other paired-data situations. Biometrics 40, 1025-35.

Scott, A.J., and Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods, Journ. Amer. Statist. Assoc. 77, 848-54.

Shirley, E.A.C., and Hickling, S.W. (1981). An evaluation of some statistical methods for analysing numbers of abnormalities found amongst litters in teratology studies. Biometrics 37, 819-29.

Siegel, D.G., and Greenhouse, S.W. (1973). Multiple relative risk functions in case-control studies. Amer. Journ. Epid. 97, 324-31.

Skellam, J.G. (1948). A probability distribution derived from the binomial distibution by regarding the probability of success as variable between the sets of trials. Journ Roy. Statist. Soc. B 10, 257-61.

Smith. P.G., and Pike, M.C. (1976). Generalisations of two tests for the detection of household aggregation of disease. Biometrics 32, 817-28.

Smith, P.G., Pike, M.C., Hill, A.P., Breslow, N.E., and Day, N.E. (1981). Multivariate conditional logistic analysis of stratum-matched case-control studies. Appl. Statist. 30: 190-7.

Snedecor, G.W., and Cochran, W.G. (1967). Statistical Methods, 6th edition. Iowa State University Press, Ames.

Stiratelli, R. Laird, N. and Ware, J. (1984). Random effects models for serial observations with binary response. Biometrics 40, 961-72.

Thomas, D.R., and Rao, J.N.K. (1984). A monte carlo study of exact levels for chi-squared goodness-of-fit statistics under cluster sampling. submitted to Journ. Amer. Statist. Assoc..

Truett, J., Cornfeld, J., and Kannel, W. (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. Journ. Chron. Dis. 20, 511-24.

Wald, A. (1949). Note on the consistency of maximum likelihood estimate. Ann. Math. Statist. 20, 595-601.

Walker, S.H., and Duncan, D.B. (1967). Estimation of the probability of an event as a function of several independent variables. Biometrika 54, 167-79.

Walsh, J.E. (1947). Concerning the effect of intraclass correlation on certain significance tests. Ann. Math. Statist. 18, 88-96.

Watson, G.S. (1955). Serial correlation in regression analysis I. Biometrika 42, 327-41.

Watson, G.S. (1967). Linear least squares regression. Ann. Math. Statist. 38, 1679-99.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments invlolving reproduction and teratogenicity. Biometrics 31, 949-52.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. Appl. Statist 31, 144-8.

Winer, B.J. (1971). Statistical Principles in Experimental Design. Mcgraw-Hill, London.

Wiorkowski, J.J. (1975). Unbalanced regression analysis with residuals having a covariance structure of intra-class form. Biometrics 31, 611-18.

END

13 04 87

FIN