Multivariate nonlinear models for vectors of proportions: a generalized least squares approach

Jorge Guillermo Morel
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/11711

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17”x 23” black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6”x 9” black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Multivariate nonlinear models for vectors of proportions: A generalized least squares approach

Morel, Jorge Guillermo, Ph.D.
Iowa State University, 1987
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages √
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received
16. Other

University Microfilms International
Multivariate nonlinear models for vectors of proportions:
A generalized least squares approach

by

Jorge Guillermo Morel

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major/Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1987
# TABLE OF CONTENTS

1. INTRODUCTION  
   1.1. Introduction  
   1.2. Quantal Response Models  
   1.3. Extra Variation Models  
   1.4. Quasi-likelihood Models  
   1.5. Logistic Models Under Complex Sample Designs  
   1.6. A General Set Up  

2. DEFINITIONS AND THEOREM  
   2.1. Matrix Operations  
   2.2. Matrix Inversion  
   2.3. Convergence Theorems  
   2.4. Orders in Probability  
   2.5. Theorems for Continuous Functions on Compact Sets  

3. NONLINEAR MODEL FOR VECTORS OF PROPORTIONS  
   3.1. The General Model  
   3.2. The Minimum Distance Estimator $\hat{\beta}$  
   3.3. Strong Consistency of $\hat{\beta}$  
   3.4. The Estimator $\hat{\beta}$  
   3.5. Asymptotic Normality of $\hat{\beta}$  
   3.6. Asymptotic Normality of $\tilde{\beta}$  
   3.7. The One Step Gauss-Newton Estimator $\hat{\beta}^{(1)}$  

4. APPLICATIONS TO LOGIT MODELS  
   4.1. A Logistic Multinomial Model
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2. A Scaled Multinomial Variance Model</td>
<td>64</td>
</tr>
<tr>
<td>4.3. A Generalized Multinomial Variance Model</td>
<td>68</td>
</tr>
<tr>
<td>4.4. An Extended Scaled Multinomial Variance Model</td>
<td>71</td>
</tr>
<tr>
<td>4.5. A Model with a Finite Number of Different Covariance Matrices</td>
<td>78</td>
</tr>
<tr>
<td>4.6. A Multinomial Variance Model under Complex Sampling Designs</td>
<td>80</td>
</tr>
<tr>
<td>4.7. An Example</td>
<td>86</td>
</tr>
<tr>
<td>5. A MONTE CARLO STUDY</td>
<td>95</td>
</tr>
<tr>
<td>5.1. Data Generation Under Two Sampling Schemes</td>
<td>95</td>
</tr>
<tr>
<td>5.2. Computed Statistics</td>
<td>100</td>
</tr>
<tr>
<td>5.3. Results</td>
<td>103</td>
</tr>
<tr>
<td>6. BIBLIOGRAPHY</td>
<td>115</td>
</tr>
<tr>
<td>7. ACKNOWLEDGMENTS</td>
<td>118</td>
</tr>
<tr>
<td>8. APPENDIX A</td>
<td>119</td>
</tr>
<tr>
<td>9. APPENDIX B</td>
<td>120</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1. Introduction

Analyses of proportions that link expectations to a reduced set of parameters via sufficiently smooth functions, traditionally have been done using quantal response models. Early developments were inspired by Bliss (1935) and Berkson (1944, 1951). These models express expectation of a binomial response as a function of some cumulative distribution function evaluated at a linear combination of an unknown set of parameters and an observed vector of covariates.

In a biological context, Crowder (1978) and Williams (1982), proposed quantal response models that fit binomial models with extra variation. In cases where the likelihood cannot be defined, McCullagh and Nelder (1983) use the quasi-likelihood approach to obtain a simple model for logistic multinomial data with extra variation. The use of logistic analysis for complex sample designs was discussed by Roberts, Rao and Kumar (1987), and Fuller et al. (1987).

This research is concerned with the development of more general methods for the analysis of observed vectors of proportions when their expectations are linked to a finite set of covariates via a sufficiently smooth function. The exact distribution or likelihood function of the observed vector of proportions need not be specified, so maximum likelihood estimation is not considered. Instead, a minimum distance estimator for the model parameters is developed which basically relies on the existence of the first two moments of the observed vectors of
proportions. Some additional regularity conditions are required in order to establish asymptotic properties of the minimum distance estimator and an associated One Step Gauss-Newton estimator. Some methods of estimation for extra variation parameters are also investigated.

An important feature of the general model proposed here is that the quantal response models, including those with extra variation and complex sample designs, can be viewed as particular cases of the general model if the form of the first two moments are properly specified.

The first chapter provides a brief review of models for analyzing observed vectors of proportions when their expectations are linked to a finite set of covariates and extra-variation is present. In addition, the more general model is briefly described. In Chapter 2, some basic definitions as well as some theorems used in subsequent chapters, are presented. Chapter 3 presents a complete description of the general model and establishes the asymptotic properties of a minimum distance estimator and an associated One Step Gauss-Newton estimator. In Chapter 4, several applications to logit models are presented. Data obtained from Indonesia in 1980 are used to illustrate some of the estimation procedures. Finally, the results of a Monte Carlo study are presented. In this study, the finite sample properties of the estimation procedures are investigated and chi-square tests involving model parameters are explored.
1.2. Quantal Response Models

Let \( y_1, y_2, \ldots, y_n \) be \( n \) independent binary responses such that

\[
y_j = 1 \quad \text{if the } j\text{-th response is a success}
\]

\[
y_j = 0 \quad \text{otherwise.} \tag{1.2.1}
\]

Define \( \pi_j \) to be the probability of observing a success at the \( j\text{-th} \) trial, that is,

\[
\text{Prob}(y_j = 1) = \pi_j \tag{1.2.2}
\]

The value of \( \pi_j \) may depend on the values of some covariates

\[ x_j = (x_{1j}, x_{2j}, \ldots, x_{kj})' \]

via some smooth function. For each trial \( j, \ j=1, 2, \ldots, n \), \((y_j, x_j)\) is observed. Assume that given \( x_j \), the conditional probability of success \( \pi_j \) can be expressed as

\[
\pi_j = F(x_j' \beta^0) \tag{1.2.3}
\]

where \( \beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_k^0)' \) is an unknown vector of parameters and \( F(\cdot) \) is a cumulative distribution function. A model that satisfies (1.2.1)-(1.2.3) is called a quantal response model.

If the cumulative distribution function (1.2.3) corresponds to the normal distribution, then the relations (1.2.1)-(1.2.3) define a probit model. In connection with bioassay, Bliss (1935), analyzed data using a probit model. He assumed that the probability \( \pi_j \) is related to some
covariate $x_j$ (dose level measured on some suitable scale) as

$$
\pi_j = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-t^2/2} dt
$$

where $\Phi(\star)$ is the cumulative standard normal distribution function. A comprehensive discussion of probit models can be found in Finney (1971).

In contrast with probit models, logit models assume that (1.2.3) is the cumulative logistic distribution function, i.e.,

$$
\pi_j = \frac{1 + \exp(x_j^\prime \beta^0)}{1 + \exp(x_j^\prime \beta^0)}
$$

Logistic models have also been used in the context of bioassay. Two early discussions were presented by Berkson (1944, 1951). Cox (1970) pointed out that when $0.1 < \pi_j < 0.9$, the two models are nearly identical. For more extreme values, the logistic density function has thicker tails. Haberman (1974) has provided a comprehensive discussion of the large sample theoretical properties of quantal response models where the observed counts have multinomial or Poisson distributions with no extra variation.
1.3. Extra Variation Models

The basic idea of these models is that in many situations the residual variation obtained after fitting the logistic multinomial model may be greater than that which can be attributed to the binomial sampling variation assumed by the model. For instance, in toxicological studies, the experimental unit is often a litter of animals and the response is the proportion of animals exhibiting a certain trait. The proportion being studied may vary among litters treated identically because of unrecorded genetic and environmental influences. Let \( t_1, t_2, \ldots, t_n \) be \( n \) independent variables such that \( t_j = m_j \pi_j \) is distributed binomial \( (m_j, \pi_j) \) where \( \pi_j \) satisfies (1.2.5). In addition to this, assume the existence of a set \( \theta_1, \theta_2, \ldots, \theta_n \) of unobservable identically distributed continuous random variables on the open interval \((0, 1)\) such that

\[
E(\theta_j) = \pi_j \tag{1.3.1}
\]

and

\[
\text{Var}(\theta_j) = \zeta^2 \pi_j (1 - \pi_j) \tag{1.3.2}
\]

where \( \zeta^2 \) is the parameter that allows for extra variation.

Conditionally on \( \theta_j \),

\[
t_j \sim \text{Binomial}(m_j, \theta_j) \tag{1.3.3}
\]
Then, it can be shown that the unconditional first two moments of $P_j = t_j/m_j$ are

$$E(P_j) = \pi_j \tag{1.3.4}$$

and

$$\text{Var}(P_j) = \left[1 + (m_j - 1)\xi^2\right]m_j^{-1}\pi_j(1 - \pi_j) \tag{1.3.5}$$

Observe that under a cluster sample design (1.3.5) resembles the cluster variance for proportions with intracluster correlation coefficient $\xi$. A derivation and discussion of the variance for proportions in terms of the intraclass correlation can be found in Cochran (1977). If no extra-variation is present, the parameter $\xi^2$ must be zero and (1.3.5) becomes the variance for binomial proportions.

Crowder (1978) proposed a method for the regression analysis of proportions based on the beta-binomial distribution. Assume that the response variable, the observed count $t_{j\ell}$, corresponds to a one-way classificatory model where $j = 1, 2, \ldots, a$, and $\ell = 1, 2, \ldots, n_j$.

Further, assume $t_{j\ell}$ has a binomial $(m_{j\ell}, \pi_{j\ell})$ distribution conditionally on $\pi_{j\ell}$ and for each treatment or group $j$, $j = 1, 2, \ldots, a$, the parameters $\pi_{j1}, \pi_{j2}, \ldots, \pi_{jn_j}$ are independent and identically distributed as beta $(\alpha_j, \beta_j)$ random variables. Thus,

$$E(\pi_{j\ell}) = \alpha_j(\alpha_j + \beta_j)^{-1} \tag{1.3.6}$$

$$= \pi_j$$
and

\[ \text{Var}(\pi_{j}) = \phi_j \pi_j (1 - \pi_j), \]  

(1.3.7)

where \( \phi_j = (1 + \alpha_j + \beta_j)^{-1} \). To facilitate estimation, Crowder further assumed homogeneity of variance between data sets, i.e., \( \phi_j = \phi \).

Under the assumption of homogeneity, the first two unconditional moments of \( \pi_{j} \) have the same form for the Crowder and Williams models with \( \phi^2 = \phi \).

1.4. Quasi-likelihood Models

In cases where the likelihood cannot be specified, Wedderburn (1974) defined a quasi-likelihood function which requires only specification of relationship between the mean and the variance of the observations. Let \( y_1, y_2, \ldots, y_n \) be independent observations such that \( \mathbb{E}(y_j) = \mu_j \) and \( \text{Var}(y_j) = \text{V}(\mu_j) \). The quasi-likelihood (or more properly, the log-quasi-likelihood) is given by the system of partial differential equations

\[ \frac{\partial \ell(y_j; y_j)}{\partial \mu_j} = \frac{y_j - \mu_j}{\text{V}(\mu_j)}, \quad j=1, 2, \ldots, n. \]  

(1.4.1)

Wedderburn (1974) proved that for the one-parameter exponential family the log-likelihood is the same as the quasi-likelihood. As an illustration he analyzed a data set obtained from a completely
randomized block experiment where the response is the percentage leaf area of barley infected with a certain disease. McCullagh (1983) extended the log-quasi-likelihood functions to multivariate random vectors and established several asymptotic properties of the quasi-likelihood estimates. McCullagh and Nelder (1983) formulate models for analyzing observed vectors of proportions where the multinomial assumption seems to be unrealistic. In such cases, the variation in the observed vector of proportions is greater than one would expect for the multinomial model. That phenomenon is often referred to as over-dispersion or heterogeneity. In some circumstances it is reasonable to assume

\[ E(P^*_j) = \pi^*_j \]  

and

\[ \text{Var}(P^*_j) = \phi_m^{-1} (\text{Diag} \pi^*_j - \pi^*_j \pi^*_j') , \]

where \( P^*_j \) is an observed vector of proportions, \( \pi^*_j \) is linked to a reduced set of covariates through a generalized logistic function, and \( \phi > 0 \) is an extra-variability parameter.

1.5. Logistic Models Under Complex Sample Designs

In the last few years a lot of attention has been given to the problems that arise when chi-square tests based on the multinomial distribution are applied to data obtained from complex sample designs. It has been shown that the effects of stratification and clustering on
the chi-square tests may lead to a distortion of nominal significance levels. Holt, Scott and Ewings (1980) proposed modified Pearson chi-square statistics tests of goodness-of-fit, homogeneity, and independence in two-way contingency tables. Rao and Scott (1981) presented similar tests for complex sample surveys. In all these cases, the correction factor requires only the knowledge of variance estimates (or design effects) for individual cells. Bedrick (1983) derived a correction factor for testing the fit of hierarchical log linear models with closed form parameter estimates. Rao and Scott (1984) presented more extensive methods of using design effects to obtain chi-square tests for complex surveys. They generalized their previous results to multi-way tables. Fay (1985) presented the adjustments to the Pearson and likelihood test statistics through a jackknife approach.

More recently, Roberts, Rao and Kumar (1987) showed how to make adjustments that take into account the survey design in computing the standard chi-square and the likelihood ratio test statistics for logistic regression analysis involving a binary response variable. The adjustments are based on certain generalized designed effects. Their results can be applied to cases where the whole population has been divided into $I$ domains of study, a large sample is obtained for each domain, and in each domain a proportion $\pi_i$, $i=1, 2, \ldots, I$, is to be estimated. It is assumed

$$\pi_i = \left[1 + \exp(x_i'|z_0)\right]^{-1}\exp(x_i'|z_0), \quad i=1, 2, \ldots, I. \quad (1.5.1)$$
This procedure may be most useful when only the summary table of counts and variance adjustment factors are available, instead of the complete data set.

Fuller et al. (1987) incorporated into PC CARP the following method for estimating the parameters of a generalized logistic model when a complex sample design is employed. If \( \hat{\theta} \) is the estimate of the parameters of the logistic model that results from assuming the multinomial distribution, then, under regularity conditions, it can be shown that

\[
\sqrt{n}(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, H^{-1}G^*H^{-1}),
\]

(1.5.2)

where \( \theta^0 \) is the true parameter vector, \( H^{-1} \) is the asymptotic covariance matrix of the parameter estimates under the multinomial assumption, and \( H^{-1}G^*H^{-1} \) is the asymptotic covariance matrix of the parameter estimates under the complex sample design. In this approach, \( \hat{\theta} \) and \( \hat{H}^{-1} \) are first computed using a modified Gauss-Newton iterative procedure. Once the algorithm converges, the asymptotic covariance matrix of the parameter estimates is re-evaluated in such a way that the different sources of variation due to the sampling design are incorporated.

1.6. A General Set Up

Let \( \mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n \) be a sequence of \( n \) vectors of proportions such that for each \( j, j=1, 2, \ldots, n \),
where \( \eta_j^* \) is a vector of means and \( \varepsilon_j^* \) a random error vector.

The general model developed in this research assumes that the error vectors \( \varepsilon_j^* \) are independent and that the elements of \( \eta_j^* \) are smooth functions of a reduced set of parameters \( \theta^0 \) and an observed vector of covariates \( x_j \), that is

\[
\eta_j^* = f(\theta^0, x_j).
\]

Furthermore, the covariance matrix for each \( \varepsilon_j^* \) is a function of \( \eta_j^* \) and possibly an extra set of parameters \( \theta^0 \), i.e.,

\[
\text{Var}(\varepsilon_j^*) = A^*_j(\eta_j^*, \theta^0)
\]

Observe that the first two moments of all models reviewed in this chapter can be obtained by properly choosing \( f \) and \( A^*_j \).

Obviously, maximum likelihood estimation is not always possible for this general model because the distribution of \( P_j^* \) need not be specified. Instead, a generalized least squares approach is used. One possibility is a minimum distance estimator \( \hat{\theta} \), defined as any value of \( \theta \) that minimizes the squared distance function.
This estimator should be invariant to the choice of the generalized inverse $A_j^*$. Since the evaluation of the minimum distance estimator generally requires an iterative procedure, a more easily evaluated One Step Gauss-Newton estimator of $\hat{\beta}^0$ is also investigated. This latter estimator is based upon the existence of consistent estimators of $\hat{\beta}^0$ and $\hat{\lambda}^0$. The asymptotic properties of these estimators are presented in detail in Chapter 3. Applications and illustrations to logistic models are shown in Chapter 4. Finally, results of a Monte Carlo study are reported in Chapter 5 to provide some information about small sample properties of the estimators and corresponding test statistics.
2. DEFINITIONS AND THEOREMS

This chapter presents some of the basic definitions, notations and theorems used in subsequent chapters. When the proofs of the theorems are omitted, references to available proofs are given.

2.1 Matrix Operations

These operations provide convenient means to express and manipulate some of the models considered in subsequent chapters. Unless otherwise specified, matrices will be denoted by \( A \) and column vectors by \( a \). A "\( - \)" will be placed beneath any greek letter used to denote matrices or vectors, i.e., \( \alpha, \beta, \gamma, \ldots \).

**Definition 2.1.1.** Kronecker Product. Let \( A = (a_{ij}) \) and \( B = (b_{ij}) \) be \( m \times n \) and \( p \times q \) matrices, respectively. Then the Kronecker product of \( A \) and \( B \), denoted by \( A \otimes B \), is the \( mp \times nq \) matrix

\[
A \otimes B = \begin{pmatrix}
    (a_{11}B, \ldots, a_{1n}B) \\
    \vdots \\
    (a_{m1}B, \ldots, a_{mn}B)
\end{pmatrix}
\]

**Definition 2.1.2.** Vec Operation. Let \( A = (a_{ij}) \) be an \( m \times n \) matrix, and let \( a_{.j} \) denote the \( j \)-th column of \( A \). Then

\[
\text{Vec } A = (a_{11}, \ldots, a_{1l}, a_{12}, \ldots, a_{1n}, a_{21}, \ldots, a_{2m}, \ldots, a_{ln}, \ldots, a_{mn})'
\]
Observe that Vec $A$ is the $mn \times 1$ vector that results from concatenating the columns of $A$, one beneath the other, in a single column vector. When the matrix $A$ is a symmetric $n \times n$ matrix, Vec $A$ will contain $n(n - 1)/2$ pairs of identical elements. In some situations, it is convenient to retain only one element of each pair. This can be achieved by listing by columns the elements in the lower triangular part of $A$.

**Definition 2.1.3.** Vech Operation. Let $A = (a_{ij})$ be an $n \times n$ symmetric matrix, then

$$\text{Vech } A = (a_{11}', a_{21}', \ldots, a_{n1}', a_{22}', a_{32}', \ldots, a_{n2}', \ldots, a_{nn}').$$

**Definition 2.1.4.** Diag Operation. Let $a = (a_1, \ldots, a_n)'$ be an $n \times 1$ vector, then Diag $a$ is an $n \times n$ matrix such that the $i$-th diagonal element is the $i$-th element of the vector $a$ and each off-diagonal element is zero. The operation can be defined on row vectors as well.

**Definition 2.1.5.** Block Vec Operation. Let $A_1, A_2, \ldots, A_n$ be $n$ matrices each of dimensions $p \times q$. Then

$$\text{Block Vec}(A_1, A_2, \ldots, A_n) = (A_1', A_2', \ldots, A_n').$$
Note that Block Vec($A_1, A_2, \ldots, A_n$) is the $np \times q$ matrix that results from concatenating the matrices $A_1, A_2, \ldots, A_n$, beneath the other.

**Definition 2.1.6. Block Diag Operation.** Let $A_1, A_2, \ldots, A_n$ be $n$ matrices each of dimension $p \times p$. Then

$$\text{Block Diag}(A_1, A_2, \ldots, A_n) = \begin{pmatrix}
A_1 & 0 & \cdots & 0 \\
0 & A_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_n
\end{pmatrix}.$$ 

where $0$ is a $p \times p$ null matrix.

**Definition 2.1.7. A Operation.** Let $a = (a_1, a_2, \ldots, a_n)'$ be an $n \times 1$ vector, then

$$A(a) = \text{Diag} \ a - a \ a'.$$

### 2.2 Matrix Inversion

Some formulas for inverting special types of matrices are now presented. These results are used in subsequent chapters to derive results involving covariance matrices estimated from multinomial and two-stage cluster samples.
Theorem 2.2.1. Let $A$ be an $n \times n$ nonsingular matrix and let $x$ be an $n \times 1$ vector. Then

$$(A - xx')^{-1} = A^{-1} + (1 - \alpha)^{-1} A^{-1} x x' A^{-1},$$

where $\alpha = x'A^{-1}x$, provided $1 - \alpha \neq 0$.


Corollary 2.2.2. Let $\pi = (\pi_1, \ldots, \pi_d)'$ be a $d \times 1$ vector such that $\pi_i > 0$ for each $i = 1, 2, \ldots, d$ and $\sum_{i=1}^d \pi_i < 1$. Then

$$[\text{Diag}(\pi) - \pi \pi']^{-1} = \text{Diag}(u) + \gamma^{-1} J,$$

where $u = (\pi_1^{-1}, \ldots, \pi_d^{-1})$, $\gamma = 1 - \sum_{i=1}^d \pi_i$, and $J = 1 1'$.  

Proof. Since $\gamma = 1 - \sum_{i=1}^d \pi_i$ is positive, $\gamma^{-1}$ is well defined. Specifying $A = \text{Diag}(\pi)$ and $x = \pi$, the result follows from Theorem 2.2.1.

2.3 Convergence Theorems

These results are used to derive asymptotic properties and limiting distributions of estimators in the next chapter. Unless otherwise specified, limits are always assumed tending to infinite. For the proof of the first result, see Theorem 5.2.1 in Fuller (1976).
Theorem 2.3.1. Let \( \{X_n\} \) and \( \{Y_n\} \) be sequences of random variables such that \( X_n - Y_n \xrightarrow{P} 0 \). If there exists a random variable \( X \) such that \( X_n \xrightarrow{d} X \), then \( Y_n \xrightarrow{d} X \).

A multivariate version of this theorem is given by the following corollary.

Corollary 2.3.2. Let \( \{x_n\} \) and \( \{y_n\} \) be sequences of \( r \)-dimensional random vectors such that \( x_n - y_n \xrightarrow{P} 0 \). If there exists a \( r \)-dimensional random vector \( x \) such that \( x_n \xrightarrow{d} x \), then \( y_n \xrightarrow{d} y \).

Proof. Let \( a \) be a \( r \)-dimensional vector of fixed constants. Since for each \( i = 1, 2, \ldots, r \), \( X_{1n} - Y_{1n} \xrightarrow{P} 0 \), then \( a'(x_n - y_n) \xrightarrow{P} 0 \). Furthermore, \( x_n \xrightarrow{d} x \), and by Theorem 2.9.4 in Billingsley (1979), \( a'x_n \xrightarrow{d} a'x \). Then, by Theorem 2.3.1, \( a'y_n \xrightarrow{d} a'y \), and by the converse of Billingsley's Theorem 2.9.4, \( y_n \xrightarrow{d} y \).

The next theorem is also in Fuller (1976).

Theorem 2.3.3. Let \( \{y_n\} \) be a sequence of \( r \)-dimensional random vectors and let \( \{A_n\} \) be a sequence of \( r \times r \) random matrices. If there exists a random vector \( y \) and a fixed nonsingular matrix \( A \) such that \( y_n \xrightarrow{d} y \) and \( A_n \xrightarrow{P} A \), then \( A_{-1}y_n \xrightarrow{d} A_{-1}y \).

Theorem 2.3.4. Let \( \{y_n\} \) be a sequence of uncorrelated \( r \)-dimensional random vectors such that \( E(y_n) = 0 \) and for each \( i = 1, 2, \ldots, r \), \( \text{Var}(Y_{1n}) \) is uniformly bounded, then
Proof. The theorem is true since each component of \( n^{-1} \sum_{j=1}^{n} y_j \) converges almost surely to zero by Theorem 5.1.2 in Chung (1974).

Theorem 2.3.5. Multivariate Central Limit Theorem. Let \( \{y_j\} \) be a sequence of independent \( r \)-dimensional random vectors such that \( E(y_j) = 0 \) and \( \text{Var}(y_j) = A_j \). Suppose that as \( n \to \infty \),

\[
n^{-1} \sum_{j=1}^{n} A_j + V \neq 0
\]

and for every \( \varepsilon > 0 \)

\[
n^{-1} \sum_{j=1}^{n} \int_{\|y\| > \varepsilon \sqrt{n}} \|y\|^2 dF_j + 0
\]

where \( F_j \) is the distribution function of \( y_j \) and \( \|y\| \) denotes the norm of \( y \). Then

\[
n^{-1/2} \sum_{j=1}^{n} y_j \xrightarrow{L^2} N_r(0, V).
\]

2.4 Orders in Probability

Let \( \{X_n\} \) be a sequence of random variables and let \( \{a_n\} \) be a sequence of real numbers.

Definition 2.4.1. \( X_n \) is at most of order in probability \( a_n \), denoted by \( X_n = O_{p}(a_n) \) if, for every \( \varepsilon > 0 \) there exists a positive real number \( M_\varepsilon \) such that

\[
\Pr\left(\frac{|X_n|}{a_n} > M_\varepsilon \right) < \varepsilon \quad \text{for all } n.
\]

Definition 2.4.2. \( X_n \) is of smaller order in probability than \( a_n \), denoted by \( X_n = o_{p}(a_n) \) if

\[
\frac{a_n}{X_n} \xrightarrow{p} 0.
\]

These two definitions can be extended to random matrices as follows.

Let \( X_n \) be a \( p \times d \) dimensional random matrix and let \( \{a_n\} \) be a sequence of real numbers.

Definition 2.4.3. \( X_n \) is at most of order in probability \( a_n \), denoted by \( X_n = O_{p}(a_n) \) if

\[
X_{ijn} = O_{p}(a_n) \quad \text{for each } i = 1, 2, \ldots, p, \text{ for each } j = 1, 2, \ldots, d.
\]

Definition 2.4.4. \( X_n \) is of smaller order in probability than \( a_n \),
denoted by $X_n = o_p(a_n)$ if

$X_{ijn} = o_p(a_n)$ for each $i = 1, 2, \ldots, p$, for each $j = 1, 2, \ldots, d$.

Useful properties of orders in probability can be found in Pratt (1959).

2.5 Theorems for Continuous Functions on Compact Sets

Two basic results are presented here which provide specific properties of continuous functions on compact sets.

**Theorem 2.5.1.** Let $g: \mathbb{R}^d \to \mathbb{R}$ be a real-valued function. If $g$ is continuous on a compact subset $C$ of $\mathbb{R}^d$, then

1) There exists a constant $M$ such that

$$|g(x)| < M \text{ for all } x \in C.$$

ii) There exist points $x_1$ and $x_2$ in $C$ such that

$$g(x_1) < g(x) < g(x_2) \text{ for all } x \in C.$$

**Proof.** See Theorems 4.27 and 4.28 in Apostol (1974).

An important implication of the first part of this theorem is that continuous functions on compact sets are uniformly bounded.

**Theorem 2.5.2.** Let $g: \mathbb{R}^d \to \mathbb{R}$ be a real-valued function such that $A < g(x) < B$ for all $x \in \mathbb{R}^d$. If $g$ is continuous on a compact subset $C$ of $\mathbb{R}^d$ then there exist constants $M_1$ and $M_2$ such that
\[ A < M_1 < g(x) < M_2 < B. \]

Proof. Follows easily from part ii) of Theorem 2.5.1.

A clear consequence of this theorem is that if \( A < g(x) < B \) then \( g(x) \) will be bounded away from \( A \) and \( B \) uniformly for all \( x \) in a compact set.
3. A NONLINEAR MODEL FOR VECTORS OF PROPORTIONS

In this chapter, a general model is developed for the analysis of independent vectors of observed proportions when the expectations and covariance matrices are linked to a reduced set of parameters via smooth, possibly nonlinear, functions. A minimum distance estimator of the reduced set of parameters is proposed. The asymptotic properties of this estimator are investigated.

3.1 The General Model

Let \( P_1^*, P_2^*, \ldots, P_n^* \) be a sequence of \((d+1)\)-dimensional independent random column vectors such that for each \( j = 1, 2, \ldots, n \),

\[
E(P_j^*) = \pi_j^*, \tag{3.1.1}
\]

and

\[
\text{Var}(P_j^*) = A_j^*, \tag{3.1.2}
\]

where \( P_j^* = (P_{1j}, \ldots, P_{dj}, P_{d+1,j})' \), \( \pi_j^* = (\pi_{1j}, \ldots, \pi_{dj}, \pi_{d+1,j})' \),

\[
0 < P_{ij} < 1, \quad 0 < \pi_{ij} < 1 \quad \text{for each} \quad i = 1, 2, \ldots, d, d+1, \]

\[
\sum_{i=1}^{d+1} P_{ij} = 1, \quad \sum_{i=1}^{d+1} \pi_{ij} = 1, \quad \text{and} \quad \text{rank}(A_j^*) = d. \]

The vector of expected proportions \( \pi_j^* \) will be assumed to be a function of the values of a \( k \)-dimensional vector of covariates, denoted by \( x_j \), and an \( r \)-dimensional vector of parameters \( \varphi^0 \). The functions that link the \( \pi_j^* \) with the true value of the unknown parameter vector,
are denoted by

\[ \pi_{ij} = g_i(\theta^0, x_j), \text{ for } i = 1, 2, \ldots, d, \]

where \( g_i: \Theta \times \mathbb{R}^k \to (0, 1) \), \( \Theta \) is a subset of \( \mathbb{R}^k \), and

\[ \pi_{d+1,j} = 1 - \sum_{i=1}^{d} \pi_{ij}. \]

For each \( i = 1, 2, \ldots, d \), \( g_i(\theta, x) \) is continuous for all \( \theta \in \Theta \), \( x \in \mathbb{R}^k \), and has continuous first, second and third partial derivatives with respect to the elements of \( \theta \).

Initially, consider the model for which each covariance matrix \( A_j^* \) is known. Later, a more practical case is considered which allows the elements of \( A_j^* \) to be functions of \( \theta^0 \) and, possibly, an additional finite set of parameters \( \xi^0 \).

3.2 The Minimum Distance Estimator \( \hat{\theta} \)

Given the observed vector of proportions, \( \mathbf{p}_1^*, \mathbf{p}_2^*, \ldots, \mathbf{p}_n^* \), an estimator for \( \theta^0 \), denoted by \( \hat{\theta}^{(n)} \), is any \( \theta \in \Theta \) that minimizes the squared distance

\[ Q_n(\theta) = n^{-1} \sum_{j=1}^{n} \left[ \mathbf{p}_j^* - \pi_j^*(\theta) \right]' \left( A_j^* \right)^{-1} \left[ \mathbf{p}_j^* - \pi_j^*(\theta) \right], \]

where

\[ \pi_j^*(\theta) = \{ g_1(\theta^0, x_j), \ldots, g_d(\theta^0, x_j), 1 - \sum_{i=1}^{d} g_i(\theta^0, x_j) \}'. \]
and $(A_j^*)^{-}$ is a generalized inverse of $A_j^*$. It will be assumed that the vector $F_j^* - \pi_j^*(\beta)$ belongs to the column space of $A_j^*$. This assumption can be verified in all the applications considered in Chapter IV. If $F_j^* - \pi_j^*(\beta)$ is perpendicular to the vector of ones $1_{(d+1)}$, then the assumption that $F_j^* - \pi_j^*(\beta)$ is in the column space of $A_j^*$ is equivalent to $A_j^* 1_{(d+1)} = 0$.

For each $j = 1, 2, ..., n$, define $F_j$ to be the $d$-dimensional random vector that remains after deleting the last component of $F_j^*$. Similarly, $\pi_j$ is derived from $\pi_j^*$ by deleting the last component.

Let $A_j$ be the matrix that remains after deleting the last row and the last column from $A_j^*$. By Lemma 2.24 in Rao and Mitra (1971), the quadratic form $[F_j^* - \pi_j^*(\beta)]'(A_j^*)^{-}[F_j^* - \pi_j^*(\beta)]$ is invariant to the choice of the generalized inverse. Therefore, the squared distance function (3.2.1) is the same as

$$Q_n(\beta) = n^{-1} \sum_{j=1}^{n} [F_j - g(\beta, x_j)]' A_j^{-1} [F_j - g(\beta, x_j)],$$

where $g(\beta, x_j) = [g_1(\beta, x_j), ..., g_d(\beta, x_j)]'$ and the estimator $\hat{\beta}^{(n)}$ is invariant to the choice of the generalized inverse.

### 3.3 Strong Consistency of $\hat{\beta}$

In this section it will be shown that $\hat{\beta}^{(n)}$ converges almost surely to the true parameter vector $\beta^0$ for the model described in (3.1.1)-(3.1.3) if appropriate restrictions are placed on $\beta$, $A_j$ and $x_j$. Let $\{x_n\}$ be a sequence of covariate vectors, $x_n \in \mathbb{R}^k$. For
each \( n \), the empirical distribution function, \( F_n(z) \), \( z \in \mathbb{R}^k \), is defined as

\[
F_n(z) = n^{-1} \sum_{j=1}^{n} \psi(x_j; z),
\]

where

\[
\psi(x; z) = 1 \text{ if } x_1 < z_1, \ldots, x_k < z_k, \text{ and } \psi(x; z) = 0 \text{ otherwise.}
\]

Then, the following conditions are used to establish asymptotic properties of \( \hat{\beta} \).

**Condition 3.1.** The parameter vector \( \beta^0 \) belongs to a compact set \( \Theta \), a subset of \( \mathbb{R}^r \).

**Condition 3.2.** Each matrix in the sequence \( \{A^{-1}_n\} \) is known and the elements of the matrices are uniformly bounded.

**Condition 3.3.** The distribution function \( F_n \) in (3.3.1) converges to a cumulative distribution function \( F \), i.e., \( \lim_{n} F_n(z) = F(z) \) for any continuity point \( z \) of \( F \).

If the \( x_j \)'s are fixed, as in the case of a designed experiment, condition (3.3) is satisfied when each combination of the levels of the factors appears in the "long run" with some specific probability. When
the \( x_j \)'s are random, it is sufficient to have a random sample \( x_1, x_2, \ldots \), from \( F \), in order to achieve condition (3.3).

Before proving the strong consistency of \( \hat{\theta} \), some preliminary results are established. The first result establishes the uniform convergence of an average of matrix products.

**Theorem 3.3.1.** Let \( H(g, x) = [h_{1m}(g, x)] \) and \( T(g, x) = [t_{k\ell}(g, x)] \) be \( p \times d \) and \( d \times q \) matrices, respectively, such that \( h_{1m}(g, x) \) and \( t_{k\ell}(g, x) \) are bounded and continuous functions on \( \Theta \times \mathbb{R}^k \) for \( i = 1, 2, \ldots, p \), for \( m, k = 1, 2, \ldots, d \), and \( \ell = 1, 2, \ldots, q \).

Then, under conditions (3.1) and (3.3)

\[
\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} H(g_j, x_j)T(g_j, x_j) \text{ converges uniformly for } g, \beta \in \Theta.
\]

**Proof.** Observe that \( H(g_j, x_j)T(g_j, x_j) = [g_{1\ell}(g_j, x_j)] \), where

\[
g_{1\ell}(g_j, x_j) = \sum_{m=1}^{d} h_{1m}(g, x_j)t_{m\ell}(g, x_j),
\]

\( \chi = \text{vec}(g, \beta) \), and \( \chi \in \Theta = \Theta \times \Theta \) a compact subset in \( \mathbb{R}^{2r} \). If it could be shown, as \( n \) goes to infinity, that

\[
\sup_{\chi \in \Theta} \left| \int g_{1\ell}(\chi, x) dF_n - \int g_{1\ell}(\chi, x) dF \right| \to 0 \quad (3.3.1.1)
\]

then \( n^{-1} \sum_{j=1}^{n} g_{1\ell}(\chi, x_j) \) would converge to \( \int g_{1\ell}(\chi, x) dF \) uniformly for all \( \chi \in \Theta \), and the theorem would be true. So, it is enough to show
Without loss of generality, drop the subscripts \( i \) and \( k \) in \( g_{ik} \).

Since \( F_n(z) \) converges to \( F(z) \) in a dense set, then by Skorohod's Theorem [see Theorem 25.6 in Billingsley (1979)] there exist random vectors \( x_n \) and \( x \) on the Lebesque space such that \( x_n \xrightarrow{a.s.} x \) and \( x_n \sim F_n, x \sim F \). Then \( \forall \eta > 0 \), there exists \( A_\eta \) such that

\[
\sup_n \Pr(|x_n| > A_\eta) < \eta \quad \text{and} \quad \Pr(|x| > A_\eta) < \eta. \tag{3.3.1.2}
\]

Since \( g: \Omega \times \mathbb{R}^k + \mathbb{R} \) is continuous and bounded, and \( \Omega \) is compact, \( \forall \eta > 0 \) and \( A_\eta < \infty \), there exists \( \delta_\eta > 0 \) such that

\[
\sup_{\chi \in \Omega} \left| g(\chi, x) - g(\chi, y) \right| < \eta \quad \tag{3.3.1.3}
\]

\[
\chi \in \Omega, \quad x, y \in [-A_\eta, A_\eta],
\]

\[
|x - y| < \delta_\eta
\]

where \( x, y \in [-A_\eta, A_\eta] \) means that each component of the vectors \( x, y \) belongs to the interval \([-A_\eta, A_\eta]\) and \( |x - y| < \delta_\eta \) means that the absolute value of each component of the vector \( x - y \) is smaller than \( \delta_\eta \). Observe also that

\[
\forall \delta_\eta > 0, \lim_{n \to \infty} \Pr(|x_n - x| > \delta_\eta) = 0. \tag{3.3.1.4}
\]
Then, making appropriate use of (3.3.1.2), (3.3.1.3) and the fact that 
g \text{ is bounded}

\[ \left| \int g(\gamma, x) d\Phi - \int g(\gamma, x) d\Phi \right| = \left| E[g(\gamma, x)] - E[g(\gamma, x)] \right| \]

\[ < E\left[ |g(\gamma, x_n) - g(\gamma, x)| : |x_n| < A_\eta, |x| < A_\eta, |x_n - x| < \delta_\eta \right] \]

\[ + 2\|g\| \left[ \Pr(|x_n| > A_\eta) + \Pr(|x| > A_\eta) + \Pr(|x_n - x| > \delta_\eta) \right] \]

\[ < \eta + 2\|g\| 2n + \Pr(|x_n - x| > \delta_\eta) . \]

Then

\[ \sup_{\gamma \in \Omega} \left| E[g(\gamma, x)] - E[g(\gamma, x)] \right| < \eta(1 + 4\|g\|) + \Pr(|x_n - x| > \delta_\eta) . \]

Then, by (3.3.1.4)

\[ \limsup_{n \to \infty} \sup_{\gamma \in \Omega} \left| E[g(\gamma, x)] - E[g(\gamma, x)] \right| < \eta(1 + 4\|g\|) . \]

The result holds since \( \eta \) is arbitrary.

\[ \square \]

Corollary 3.3.1.1. Let \( H(g, x) = [h_{1m}(g, x)] \) and \( F(\beta, x) = [f_{k\beta}(\beta, x)] \) be \( p \times d \) and \( d \times q \) matrices respectively, such that \( h_{1m}(g, x) \) and \( f_{k\beta}(\beta, x) \) are bounded and continuous functions on \( \mathbb{Z} \times \mathbb{R}^k \) for \( i = 1, 2, \ldots, p \), for \( m, k = 1, 2, \ldots, d \)
and for \( r = 1, 2, \ldots, g \). Then, under conditions (3.1)-(3.3),

\[
    n^{-1} \sum_{j=1}^{n} H(\beta, x_j) A_j^{-1} F(\beta, x_j)
\]

converges uniformly for \( \alpha, \beta \in \mathcal{Q} \).

**Proof.** For each \( j = 1, 2, \ldots, n \), the elements of \( A_j^{-1} \) are known and uniformly bounded, then each element of the function

\[
    T(\beta, x_j) = A_j^{-1} F(\beta, x_j)
\]

is bounded and continuous on \( \mathcal{Q} \times \mathbb{R}^k \).

Therefore, by Theorem 3.3.1,

\[
    n^{-1} \sum_{j=1}^{n} H(\beta, x_j) T(\beta, x_j)
\]

converges uniformly for all \( \alpha, \beta \in \mathcal{Q} \).

---

**Theorem 3.3.2.** Let the model (3.1.1)-(3.1.3) hold, and let conditions (3.1)-(3.3) hold. Let \( F(\beta, x) = [f_{im}(\beta, x)] \) be a \( p \times d \) matrix such that \( f_{im}(\beta, x) \) is bounded and continuous on \( \mathcal{Q} \times \mathbb{R}^k \) for each \( l = 1, 2, \ldots, p \), \( m = 1, 2, \ldots, d \). Let \( e_j = p_j - g(\beta^0, x_j) \) for \( j = 1, 2, \ldots, n \). Then

\[
    n^{-1} \sum_{j=1}^{n} F(\beta, x_j) A_j^{-1} e_j \xrightarrow{a.s.} 0 \text{ uniformly for } \beta \in \mathcal{Q},
\]

where \( 0 \) is the \( p \times 1 \) null vector.
Proof. Let $F^X_j$ be the $\lambda$-th row of $F(g, x_j)$, i.e.,

$F^X_j = [F_{g1}(g, x_j), \ldots, F_{g\lambda}(g, x_j)]$. It is sufficient to show

$$n^{-1} \sum_{j=1}^{n} F^X_j A_j^{-1} e_j \xrightarrow{a.s.} 0$$

uniformly for $g \in \mathcal{G}$ for each $\lambda = 1, 2, \ldots, d$.

For each $j = 1, 2, \ldots, n$, define

$$W_j = e_j A_j^{-1} e_j - d$$

and for each $\lambda = 1, 2, \ldots, d$, let

$$Z_{\lambda j}(g) = F^X_j A_j^{-1} e_j, \quad g \in \mathcal{G}.$$ 

Then, by standard results in quadratic forms

$$E(W_j) = [E(e_j)]' A_j^{-1} E(e_j) = 0 \quad (3.3.2.1)$$

and

$$\text{Var}(W_j) = [\text{vec}(A_j^{-1})]' R_j [\text{vec}(A_j^{-1})] + 2d \quad (3.3.2.2)$$

where $R_j$ is a $d^2 \times d^2$ matrix for which the entry in which $kl$-th row and $mn$-th column is

$$r_{kl, mn}^{j} = \delta_{kl}^{j} - a_{kl}^{j} a_{mn}^{j} - a_{km}^{j} a_{ln}^{j} - a_{kn}^{j} a_{lm}^{j},$$
where
\[
\begin{align*}
c_{k\ell,mn}^j &= E[(P_{kj} - \pi_{kj})(P_{\ell j} - \pi_{\ell j})(P_{mj} - \pi_{mj})(P_{o j} - \pi_{o j})] \\
& \quad \text{for } k, \ell, m, o = 1, 2, \ldots, d,
\end{align*}
\]
and \(a_{k\ell}^j\) is the \((k, \ell)\)-th element of \(A_j\).

Since for all \(j = 1, 2, \ldots, n\) and for all \(i = 1, 2, \ldots, d\),
\(0 < P_{ij} < 1\) with probability one, it follows that the first four
moments are bounded uniformly. Consequently, there exists a constant
\(M\) such that
\[
|c_{k\ell,mn}^j| < M. \quad (3.3.2.3)
\]

Then, by condition (3.2), \(Var(W_j)\) is uniformly bounded for all
\(j = 1, 2, \ldots, n\), and by (3.3.2.1) and Theorem 2.3.4,
\[
n^{-1} \sum_{j=1}^{n} W_j \xrightarrow{a.s.} 0. \quad \text{Therefore,}
\]
\[
n^{-1} \sum_{j=1}^{n} e_j^{A_j^{-1}} e_j \xrightarrow{a.s.} d. \quad (3.3.2.4)
\]

Similarly,
\[
E[Z_{x j}(\alpha)] = 0
\]
and
\[
Var[Z_{x j}(\alpha)] = F_{\alpha}(\alpha_j, x_j)A_j^{-1}[F_{\alpha}(\alpha_j, x_j)]'.
\]
is uniformly bounded. Then by Theorem 2.3.4,

\[ n^{-1} \sum_{j=1}^{n} Z_{\alpha,j}(g) \xrightarrow{a.s.}\ 0 \text{ for each } g \in \mathcal{G}. \]  

(3.3.2.5)

By Corollary 3.3.1.1,

\[ n^{-1} \sum_{j=1}^{n} [F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]A_{j}^{-1}[F_{\alpha}(g, x) - F_{\beta}(g, x)]' \]

converges uniformly in \( \alpha, \beta \in \mathcal{G} \). Since \( F_{\alpha} \) is continuous, for every \( n > 0 \) and every \( g \in \mathcal{G} \), there exists a neighborhood \( N(g, \eta) \), and an integer \( i(g, \eta) \) such that \( n > i(g, \eta) \) implies

\[ n^{-1} \sum_{j=1}^{n} [F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]A_{j}^{-1}[F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]' < \eta \]

(3.3.2.6)

for all \( \beta \in N(g, \eta) \). Now observe that

\[
|n^{-1} \sum_{j=1}^{n} F_{\alpha}(g, x_j)A_{j}^{-1}e_j| < |n^{-1} \sum_{j=1}^{n} [F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]A_{j}^{-1}e_j| \\
+ |n^{-1} \sum_{j=1}^{n} F_{\beta}(g, x_j)A_{j}^{-1}e_j| \\
< (n^{-1} \sum_{j=1}^{n} [F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]A_{j}^{-1}[F_{\alpha}(g, x_j) - F_{\beta}(g, x_j)]')^{1/2} \\
\cdot \left[ n^{-1} \sum_{j=1}^{n} e_j A_{j}^{-1}e_j \right]^{1/2} + |n^{-1} \sum_{j=1}^{n} F_{\beta}(g, x_j)A_{j}^{-1}e_j| \]  

(3.3.2.7)
by Cauchy-Schwarz inequality. By (3.3.2.4) for any \( \eta > 0 \) there exists an integer \( k(\eta) \) such that \( n^{-1} \sum_{j=1}^{n} A^{-1}_{j} e_j < d + \eta \) when \( n > k(\eta) \). If \( n > i^*(g, \eta) = \max[i(g, \eta), k(\eta)] \), then by (3.3.2.6) and (3.3.2.7), for each \( g \in \Omega \) and \( \eta > 0 \),

\[
|n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j| < n^{1/2} (d + \eta)^{1/2} + |n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j|
\]

(3.3.2.8)

when \( g \in \mathcal{N}(g, \eta) \). By (3.3.2.5), \( n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j \) converges almost surely to zero vector for each \( g \in \Omega \), then the convergence of \( n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j \) takes place uniformly for \( g \in \mathcal{N}(g, \eta) \), for each \( g \in \Omega \). By (3.3.2.8) it follows that, almost surely, for each \( g \in \Omega \) and for every \( \eta > 0 \), there exist a neighborhood of \( g \), \( \mathcal{N}(g, \eta) \), and an integer \( i^*(g, \eta) \), such that if \( n > i^*(g, \eta) \)

\[
|n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j| < n^{1/2} (d + \eta)^{1/2} + \eta
\]

(3.3.2.9)

for all \( g \in \mathcal{N}(g, \eta) \). Since the union of the \( \mathcal{N}(g, \eta) \)'s, \( g \) in \( \Omega \), covers \( \Omega \), condition (3.1) implies that there exists a finite collection of neighborhoods \( \mathcal{N}(g^1, \eta), \ldots, \mathcal{N}(g^q, \eta) \) that covers \( \Omega \). Take \( n^*(\eta) = \max[i^*(g^1, \eta), \ldots, i^*(g^q, \eta)] \), then by (3.3.2.9)

\[
|n^{-1} \sum_{j=1}^{n} F_{g, x_j} A^{-1}_{j} e_j| < n^{1/2} (d + \eta)^{1/2} + \eta
\]

almost surely and uniformly in \( g \in \Omega \), whenever \( n > n^*(\eta) \).
Therefore, since \( n \) is arbitrary,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{F}_j(x_j, \omega) \mathbf{A}_j^{-1} \mathbf{e}_j \xrightarrow{a.s.} 0 \quad \text{uniformly in } \omega \in \Omega.
\]

The next result establishes the strong consistency of \( \hat{\theta}^{(n)} \) when the following identifiability condition is satisfied.

**Condition 3.4.** For any \( \alpha, \beta \in \Omega, \)
\[
\lambda(\alpha, \beta) = 0 \iff \alpha = \beta,
\]
where
\[
\lambda(\alpha, \beta) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} [g(\alpha, x_j) - g(\beta, x_j)] \mathbf{A}_j^{-1} [g(\alpha, x_j) - g(\beta, x_j)].
\]

**Theorem 3.3.3.** Let the model (3.1.1)-(3.1.3) hold, and let conditions (3.1)-(3.4) hold. Then
\[
\hat{\theta}^{(n)} \xrightarrow{a.s.} \theta^0,
\]
where \( \hat{\theta}^{(n)} \) is defined in Section 3.2.

**Proof.** Let \( (\Omega_j, \mu_j) \) be the probability space associated with the random vector \( P_j \). Define \( y = \text{vec}(P_1, \ldots, P_n) \). Since the \( P_j \)'s are independent, \( y \) is in the product probability space \( (\Omega, \nu) = \prod_{j=1}^{n} (\Omega_j, \mu_j) \). Observe that \( Q_n(\beta) \equiv Q(\beta, y) \) in (3.2.2), is a real function on \( \Theta \times \Omega \) where \( \Theta \) is a compact subset of \( \mathbb{R}^r \). Also,
note that for each $y \in \mathcal{Q}$, $Q(\beta, y)$ is a measurable function of $y$ and, for each $y$ in $\mathcal{Q}$, $Q(\beta, y)$ is a continuous function on $\mathcal{Q}$.

Then, by Lemma 2 of Jennrich (1969), there exists a measurable function $\tilde{Q}(\beta)$ from $\mathcal{Q}$ into $\mathcal{Q}$ such that for all $y \in \mathcal{Q}$,

$$ Q_n(\tilde{\beta}(n), y) = \inf_{\beta \in \mathcal{Q}} [Q_n(\beta, y)] , \text{ for each } n = 1, 2, \ldots . $$

Since the sequence $\{\tilde{\beta}(n)\}$ is in the compact set $\mathcal{Q}$, the existence of a limit point is guaranteed by the Bolzano-Weierstrass Theorem. Call this limit point $\tilde{\beta}^*$.

Now, observe that the function (3.2.1) can be written as

$$ Q_n(\beta) = Q_n(\beta^0) + 2n^{-1} \sum_{j=1}^{n} [g(\beta, x_j) - g(\beta^0, x_j)]'A_j^{-1}e_j $$

$$ + n^{-1} \sum_{j=1}^{n} [g(\beta, x_j) - g(\beta^0, x_j)]'A_j^{-1}[g(\beta, x_j) - g(\beta^0, x_j)]. $$

By Corollary 3.3.1.1, the last term on the right side of (3.3.3.1) converges to $\lambda(\beta, \beta^0)$ uniformly in $\beta \in \mathcal{Q}$. The sum of the remaining two terms converges almost surely by (3.3.2.4) and Theorem 3.3.2 to $d$, uniformly in $\beta \in \mathcal{Q}$. Therefore,

$$ Q_n(\beta) \overset{a.s.}{\longrightarrow} d + \lambda(\beta, \beta^0) , \text{ uniformly in } \beta \in \mathcal{Q} . \quad (3.3.3.2) $$

Let $\{\tilde{\beta}(n)\}$ be any subsequence of $\{\tilde{\beta}(n)\}$ which converges to the limit point $\tilde{\beta}^*$. Since $\lambda(\beta, \beta^0)$ is the uniform limit of continuous
functions, \( \lambda(\bar{g}, \bar{g}^0) \) is continuous. Then, by (3.3.3.2) and by the
continuity of \( \lambda(\bar{g}, \bar{g}^0) \), it follows that
\[
Q_n(\tilde{g}^{(n)}) \xrightarrow{a.s.} d + \lambda(\bar{g}^*, \bar{g}^0). \tag{3.3.3.3}
\]
Since \( \tilde{g}^{(n)} \) minimizes \( Q_n(\bar{g}) \), the inequality
\[
Q_n(\tilde{g}^{(n)}) < Q_n(\bar{g}^0)
\]
holds. Then, applying limits to both sides of this inequality, and
making use of (3.3.3.3) it follows that
\[
\lambda(\bar{g}^*, \bar{g}^0) < \lambda(\bar{g}^0, \bar{g}^0) \text{ almost surely.} \tag{3.3.3.4}
\]
By condition 3.4, \( \lambda(\bar{g}^0, \bar{g}^0) = 0 \). Then, by (3.3.3.4) and since
\( \lambda(\bar{g}, \bar{g}) > 0 \),
\[
\lambda(\bar{g}^*, \bar{g}^0) = 0 \text{ almost surely,} \tag{3.3.3.5}
\]
and it follows from condition (3.4)
\[
\bar{g}^* = \bar{g}^0 \text{ almost surely.}
\]
Therefore, since \( \{\tilde{g}^{(n)}\} \) is arbitrary,
3.4 The Estimator $\hat{\beta}$

Let $F(\beta, x_j)$ denote the $d \times r$ matrix of first partial derivatives of $g(\beta, x)$ with respect to the elements of $\beta$, evaluated at $\beta = \hat{\beta}$ and $x = x_j$. Consider the following linear model in $\beta$,

$$e_j = F(\beta^0, x_j) (\beta - \beta^0) + u_j$$  \hspace{5em} (3.4.1)

where $e_j = F_j - g(\beta^0, x_j)$, $E(u_j) = 0$ and $\text{Var}(u_j) = A_j$.

The estimator $\hat{\beta}^{(n)}$ of $\beta^0$, $(\hat{\beta}^{(n)} = \hat{\beta})$, is constructed on the basis of the linear model (3.4.1) by minimizing the squared distance function

$$\hat{\delta}_n(\beta) = n^{-1} \sum_{j=1}^{n} [e_j - F(\beta^0, x_j)(\beta - \beta^0)]' A_j^{-1} [e_j - F(\beta^0, x_j)(\beta - \beta^0)]$$  \hspace{5em} (3.4.2)

for $\beta \in \Theta$.

Observe that $\hat{\beta}^{(n)}$ is not a true estimator since $e_j$, $F(\beta^0, x_j)$ and $\beta^0$ are not known, so that $\hat{\beta}^{(n)}$ cannot be calculated from the observed data $(F_j, x_j)$, $j = 1, 2, \ldots, n$. But $\hat{\beta}^{(n)}$ is useful in that the asymptotic normality of $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ can be demonstrated by proving the asymptotic normality of $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ and then showing that the random vectors $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ and $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$, are
asymptotically equivalent. The next two sections deal with these properties.

3.5 Asymptotic Normality of $\hat{\beta}$

It will be shown next that $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ converges in law to a multivariate normal distribution.

Let $X^{(n)}$ be the $n \times k$ matrix of observed covariates, i.e.,

$X^{(n)} = \text{Block vec}(x_1', \ldots, x_n')$, where $x_j = (x_{j1}, \ldots, x_{jk})'$, for $j = 1, 2, \ldots, n$.

Condition 3.5. For each $n = 1, 2, \ldots$, the elements of $X^{(n)}$ belong to a compact set of $\mathbb{R}^1$.

Condition 3.6. The parameter vector $\beta^0$ belongs to the interior of $\Theta$.

Now, define

$$V_n(\beta) = n^{-1} \sum_{j=1}^{n} F'(\beta, x_j) A_j^{-1} F(\beta, x_j).$$

The elements of $F'(\beta, x)$ are continuous functions with $\beta$ and $x$ confined to compact sets. Then by Theorem 2.5.1, the elements of the sequence $\{F(\beta, x_j)\}$ are uniformly bounded. Therefore, by Corollary 3.3.1.1, $V_n(\beta)$ converges uniformly in $\beta \in \Theta$. The limit is denoted by $W(\beta)$. 
Condition 3.7. There exists a neighborhood of \( g^0 \), denoted by \( N(g^0) \), such that \( V(g) \) is nonsingular \( \forall g \in N(g^0) \).

The next theorem establishes the asymptotic distribution of \( \sqrt{n}(\hat{g}^{(n)} - g^0) \).

**Theorem 3.5.1.** Let the model (3.4.1) and let conditions (3.1)-(3.7) hold. Then

\[
\sqrt{n}(\hat{g}^{(n)} - g^0) \rightarrow N_r[0, V^{-1}(g^0)].
\]

**Proof.** Differentiation of \( \hat{q}_n(g) \) in (3.4.2) with respect to \( g \) leads to the linear system of equations

\[
0 = n^{-1} \sum_{j=1}^{n} F'(g^0, x_j) A_j^{-1} e_j - V_n(g^0)(\hat{g} - g^0), \tag{3.5.1.1}
\]

where \( V_n(g^0) = n^{-1} \sum_{j=1}^{n} F'(g^0, x_j) A_j^{-1} F(g^0, x_j) \). Then, if \( \hat{g}^{(n)} \) is a solution to (3.4.2), for sufficiently large \( n \), Condition (3.7) implies \( V_n(g^0) \) is nonsingular, and

\[
(\hat{g}^{(n)} - g^0) = V_n^{-1}(g^0)(n^{-1} \sum_{j=1}^{n} y_j), \tag{3.5.1.2}
\]

where \( y_j = F'(g^0, x_j) A_j^{-1} e_j \). By Condition (3.7),

\[
\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \text{Var}(y_j) = \lim_{n \to \infty} V_n(g^0)
\]
which is nonsingular.

The \( y_j \)'s are uniformly bounded since the elements of each vector \( e_j \) are bounded below by \(-1\) and above by \(1\), and the elements of the \( \{A_j^{-1}\} \) are uniformly bounded by condition (3.2), and the elements of \( \{F(g^0, x_j)\} \) are also uniformly bounded. Let \( F_j \) denote the distribution function of \( y_j \) and let \( \|y\| \) denote the Euclidean norm of \( y \). Then, there exists a constant \( M \) such that for all \( n \),

\[
\|y_j\|^2 < M \text{ for all } j = 1, 2, \ldots, n . \tag{3.5.1.4}
\]

Consequently,

\[
\int_{\|y\| > \varepsilon\sqrt{n}} \|y\|^2 \, dF_j < M \int_{\|y\| > \varepsilon\sqrt{n}} \, dF_j
\]

\[
= M \cdot \Pr(\|y_j\| > \varepsilon\sqrt{n})
\]

\[
< M(\varepsilon^2 n)^{-1} \mathbb{E}\|y_j\|^2 \text{ by the Chebyshev's inequality}
\]

\[
< M^2(\varepsilon^2 n)^{-1} .
\]

Then

\[
0 < \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \int_{\|y\| > \varepsilon\sqrt{n}} \, dF_j < \lim_{n \to \infty} M^2(\varepsilon^2 n)^{-1} = 0 . \tag{3.5.1.5}
\]
Since $E(y_j) = 0$, it follows from Theorem 2.3.5 that $\sqrt{n(n^{-1} \sum_{j=1}^{n} y_j)}$ converges in law to a $r$-multivariate normal with zero mean vector and covariance matrix $\mathbf{V}(\hat{\beta}^0)$. The final result follows from Theorem 2.3.3.

3.6 Asymptotic Normality of $\tilde{\beta}$

The asymptotic normality of $\tilde{\beta}$ is established by showing that $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ and $\sqrt{n}(\tilde{\beta}^{(n)} - \beta^0)$ have the same limiting distribution.

Theorem 3.6.1. Consider the models described in (3.1.1)-(3.1.3), and (3.4.1). Assume that conditions (3.1)-(3.7) are satisfied. Then

$$\sqrt{n}(\tilde{\beta}^{(n)} - \beta^0) \xrightarrow{D} N_r(0, \mathbf{V}^{-1}(\beta^0)),$$

where $\mathbf{V}(\beta^0) = \lim_{n \to \infty} \mathbf{V}_n(\beta^0)$, and

$$\mathbf{V}_n(\beta^0) = n^{-1} \sum_{j=1}^{n} \mathbf{F}'(\beta^0, x_j) \mathbf{A}_j^{-1} \mathbf{F}(\beta^0, x_j).$$

Proof. Since $\tilde{\beta}^{(n)}$ converges almost surely to $\beta^0$ which is in the interior of $\Theta$ by Condition 3.6, then the probability that $\tilde{\beta}^{(n)}$ is in the interior of $\Theta$ approaches to one as $n$ increases. The estimator $\tilde{\beta}^{(n)}$ minimizes $Q_n(\beta)$ in (3.2.2), so for large $n$ it must satisfy the necessary conditions.
\[
\frac{\partial Q_n(\vec{g})}{\partial \vec{g}} = 0 ,
\]

which can also be written as

\[
\begin{align*}
 n^{-1} \sum_{j=1}^{n} F'(\vec{g}_j, x_j) A_j^{-1} [p_j - g(\vec{g}_j, x_j)] &= 0 . \\
(3.6.1.1)
\end{align*}
\]

For each \(i = 1, 2, \ldots, d\), let \(g_1^{(i)}(\vec{g}, x)\) denote the first partial derivative of \(g_1(\vec{g}, x)\) with respect to \(\vec{g}_i\), \(i = 1, 2, \ldots, r\) and, let \(g_1^{(i)}(\vec{g}_j, x_j)\) denote \(g_1^{(i)}(\vec{g}, x)\) evaluated at \(\vec{g} = \vec{g}_j\), \(x = x_j\). Likewise, \(g_1^{(im)}(\vec{g}_j, x_j)\) is used to denote the second partial derivative of \(g_1(\vec{g}, x)\) with respect to \(\vec{g}_i\) and \(\vec{g}_m\), for \(i, m = 1, 2, \ldots, r\), evaluated at \(\vec{g} = \vec{g}_j\), \(x = x_j\). A Taylor's series expansion of (3.6.1.1) around \(\vec{g}^0\) gives for each \(i = 1, 2, \ldots, r\), and some \(\vec{q}\) on the line segment joining \(\vec{g}\) and \(\vec{g}^0\),

\[
0 = n^{-1} \sum_{j=1}^{n} h_{i,1}^{(i)}(\vec{g}_j, x_j) + (n^{-1} \sum_{j=1}^{n} A_j^{-1} [p_j - g(\vec{g}_j, x_j)])' S_{i,1}^{(i)}(\vec{g}_j, x_j) \\
- n^{-1} \sum_{j=1}^{n} u_{i,1}^{(i)}(\vec{g}_j, x_j)(\vec{g}^{(0)} - \vec{g}^0) \\
(3.6.1.2)
\]

where \(h_{i,1}^{(i)}(\vec{g}_j, x_j)\) is the \(i\)-th element of \(F'(\vec{g}_0, x_j) A_j^{-1} [p_j - g(\vec{g}_0, x_j)]\), \(F(\vec{g}_0, x_j) = [g_1^{(i)}(\vec{g}_0, x_j)]_{dxr}\), \(u_{i,1}^{(i)}(\vec{g}_j, x_j)\) is the \(i\)-th row of \(F'(\vec{g}_j, x_j) A_j^{-1} F(\vec{g}_j, x_j)\), and
Then

\[
\mathbf{0} = n^{-1} \sum_{j=1}^{n} F'(g^0, x_j) \mathbf{A}_j^{-1} e_j + G_n(g)(g^{(n)} - g^0) \quad (3.6.1.3)
\]

where

\[
G_n(g) = n^{-1} \sum_{j=1}^{n} \left( I_n \circ \{ A_j^{-1}(P_j - g(g, x_j)) \} \right)' S(g, x_j)
\]

\[- n^{-1} \sum_{j=1}^{n} F'(g, x_j) \mathbf{A}_j^{-1} F(g, x_j)\]

and

\[
S(g, x_j) = \text{Block vec}[S_1(g, x_j), \ldots, S_r(g, x_j)].
\]

Observe that

\[
n^{-1} \sum_{j=1}^{n} \left( I_r \circ \{ A_j^{-1}(P_j - g(g, x_j)) \} \right)' S(g, x_j)
\]

\[= n^{-1} \sum_{j=1}^{n} \left( I_r \circ \{ A_j^{-1} e_j \} \right)' S(g, x_j)
\]

\[+ n^{-1} \sum_{j=1}^{n} \left( I_r \circ \{ A_j^{-1}(g^0, x_j) - g(g, x_j) \} \right)' S(g, x_j) \quad (3.6.1.4)
\]
Then, by Theorem 3.3.2 and Corollary 3.3.1.1, the two terms on the right side of (3.6.1.4) both converge almost surely to a null matrix. By Condition 3.7 and Corollary 3.3.1.1,

\[ n^{-1} \sum_{j=1}^{n} F'((g, x_j) \Lambda_j^{-1} F(g, x_j) \]

converges almost surely to a nonsingular matrix \( V(g^0) \). Therefore,

\[ G_n((g) \text{ converges almost surely to } - V(g^0). \quad (3.6.1.5) \]

Since \( \tilde{g}^{(n)} \) satisfies (3.5.1.1),

\[ 0 = n^{-1} \mathbb{E}'(g^0, x_j) \Lambda_j^{-1} e_j - V_n(g^0)(\tilde{g}^{(n)} - g^0). \quad (3.6.1.6) \]

Then, by (3.6.1.3)

\[ G_n((g) (\tilde{g}^{(n)} - g^0) = - V_n(g^0)(\tilde{g}^{(n)} - g^0), \quad (3.6.1.7) \]

and by (3.6.1.5), for \( n \) sufficiently large,

\[ \sqrt{n}(\tilde{g}^{(n)} - g^0) - \sqrt{n}(\tilde{g}^{(n)} - g^0) = - \left[ G^{-1}_n((g) - V_n(g^0) + I_r \right] \sqrt{n}(\tilde{g}^{(n)} - g^0). \quad (3.6.1.8) \]

The matrix \( [G^{-1}_n((g) V_n(g^0) + I_r] \) converges almost surely by (3.6.1.5) and Condition 3.7 to a null \( r \times r \) matrix, while \( \sqrt{n}(\tilde{g}^{(n)} - g^0) \)
converges in law to a multivariate normal random vector by Theorem 3.5.1. Therefore, the right hand side of (3.6.1.8) converges in law to an r-dimensional null vector. Thus, by Corollary 2.3.2 and Theorem 3.5.1

\[ \sqrt{n}(\tilde{g}^{(n)} - g^0) \overset{d}{\to} N_r(0, \mathbf{V}^{-1}(g^0)). \]

3.7 The One Step Gauss-Newton Estimator \( \bar{g}^{(1)} \)

In Theorem 3.6.1 it was shown that \( \sqrt{n}(\tilde{g}^{(n)} - \tilde{g}^{(n)}) \) converges in probability to a null vector, implying \( \tilde{g}^{(n)} = \tilde{g}^{(n)} + o_p(n^{-1/2}) \). By (3.5.1.2)

\[ \bar{g}^{(n)} = g^0 + \mathbf{V}_n^{-1}(g^0) \left\{ \frac{1}{n} \sum_{j=1}^{n} F'(g^0, x_j) A_j^{-1} (p_j - g(g^0, x_j)) \right\}. \quad (3.7.1) \]

Then, \( \bar{g}^{(n)} \) is not a true estimator since it depends on \( \tilde{g}^0 \), the parameter vector to be estimated. Also, it is more realistic to assume \( A_j \) unknown. Since each response vector to be analyzed is a random vector of proportions satisfying the constraints that each proportion is nonnegative and the sum of the proportions is one, it is sometimes reasonable to assume that the correlation between two proportions is negative and that the covariance matrix of the observed vector of proportions \( P \), depends on \( g^0 \) via \( \mathbf{V} = \text{E}(p) \). In the more general situation, \( A_j \) may be a function of \( (g^0, x_j) \) and an additional \( u \)-dimensional parameter vector \( \phi^0 \). In Chapter 4, some applications with
the extra parameter $\phi^0$ are presented. The reason for introducing $\phi^0$ is to account for situations where the residual variation in model (3.1.1)-(3.1.3) exceeds that which can be attributed to the model when the underlying distribution is multinomial.

Let $\chi^0 = \text{Vec}(\beta^0, \phi^0)$ where $\beta^0$ is the parameter vector of interest and $\phi^0$ is the additional parameter vector associated with the extra variation. Suppose $A_j$ is a function of $(\chi^0, x_j)$, i.e.,

$$A_j = A(\chi^0, x_j) = A(\beta^0, \phi^0, x_j).$$

Then (3.7.1) generalizes to

$$\hat{\beta}^{(n)} = \beta^0 + \mathbf{v}_n^{-1}(\chi^0) \left\{ n^{-1} \sum_{j=1}^{n} F'(\beta^0, x_j) \Delta^{-1}(\chi^0, x_j) \left[ P_j - g(\beta^0, x_j) \right] \right\}$$

(3.7.2)

where

$$\mathbf{v}_n(\chi^0) = n^{-1} \sum_{j=1}^{n} F'(\beta^0, x_j) \Delta^{-1}(\chi^0, x_j) F(\beta^0, x_j).$$

If $\hat{\chi} = \text{vec}(\hat{\beta}, \hat{\phi})$ is a consistent estimator of $\chi^0$, i.e.,

$$\hat{\chi} = \chi^0 + o_p(a_n)$$

where $\lim_{n \to \infty} a_n = 0$, then the One Step Gauss-Newton (OSGN) estimator of $\beta^0$ is defined as

$$\hat{\beta}^{(1,n)} = \hat{\beta} + \mathbf{v}_n^{-1}(\hat{\chi}) \left\{ n^{-1} \sum_{j=1}^{n} F'(\hat{\beta}, x_j) \Delta^{-1}(\hat{\chi}, x_j) \left[ P_j - g(\hat{\beta}, x_j) \right] \right\}.$$

(3.7.3)

The following conditions are used to establish asymptotic properties of the OSGN-estimator.
Condition 3.8. The extra parameter vector $\xi^0$ belongs to a compact set $\xi$, a subset of $\mathbb{R}^u$.

Condition 3.9. Let $\gamma = \text{vec}(\xi^0, \xi^1)$, $\xi^0 \in \xi$, $\xi^1 \in \xi$. The elements of $A^{-1}(\gamma, x_j)$ are continuous functions of $\gamma$ with continuous first and second derivatives with respect to $\gamma$ for all $x_j$.

Condition 3.10. The extra parameter vector $\xi^0$ belongs to the interior of $\xi$.

The elements of $F(\xi, x)$ and $A^{-1}(\gamma, x)$ are continuous functions with their arguments confined to compact sets. So, by Theorem 2.5.1 and Corollary 3.3.1.1, $V_n(\gamma)$ converges uniformly for all $\gamma \in \xi \times \xi$ to some matrix denoted by $V(\gamma)$. The following condition, which is an extension of Condition 3.7, is used to establish the existence of the covariance matrix of the limiting distribution of the OSGN-estimator.

Condition 3.11. For some neighborhood $N(\gamma^0)$ of $\gamma^0$

$N(\gamma^0) \equiv N(\xi^0) \times N(\xi^1)$ where $N(\xi^0)$ and $N(\xi^1)$ are neighborhoods of $\xi^0$ and $\xi^1$ respectively, the matrix $V(\gamma)$ is nonsingular for all $\gamma \in N(\gamma^0)$.

Theorem 3.7.1. Let $\hat{\gamma}^{(1,n)}$ be the OSGN estimator defined in (3.7.3) and let $\hat{\gamma} = \text{vec}(\hat{\xi}, \hat{\xi})$ be a consistent estimator of $\gamma^0 = \text{vec}(\xi^0, \xi^1)$ such that $\hat{\gamma} = \gamma^0 + O_p(a_n)$ where $\lim \frac{a_n}{n} = 0$. Consider the model defined in (3.1.1)-(3.1.3) and the linear model defined in (3.4.1). Let conditions (3.1)-(3.11) hold. Then
\[ \hat{g}^{(1,n)} - g^0 = \sum_{j=1}^{n} F'(g^0, x_j) A^{-1}(x^0, x_j) e_j \]

\[ + O_p\left(\max(n^{-1/2} a_n, a_n^2)\right). \]

**Proof.** A Taylor's series expansion of \( g_j(g^0, x_j) \) about \( \hat{g} \) gives

\[ g_j(g^0, x_j) = g_j(\hat{g}, x_j) + [g_j^{(1)}(\hat{g}, x_j), ..., g_j^{(r)}(\hat{g}, x_j)](g^0 - \hat{g}) \]

\[ + \frac{1}{2} (g^0 - \hat{g})' H_j(\hat{g}, x_j)(g^0 - \hat{g}) \quad (3.7.1.1) \]

where \( H_j(\hat{g}, x_j) = [g_j^{(1,m)}(\hat{g}, x_j)] \) is the \( r \times r \) matrix of second partial derivatives of \( g_j(\hat{g}, x) \) evaluated at \( \hat{g} = \hat{g}, x = x_j \), and \( \hat{g} \) is on the line segment that joins \( g^0 \) and \( \hat{g} \). Then

\[ p_j - g(\hat{g}, x_j) = [g(g^0, x_j) - g(\hat{g}, x_j)] + e_j \]

\[ = F(\hat{g}, x_j)(g^0 - \hat{g}) + r(\hat{g}, x_j) + e_j \quad (3.7.1.2) \]

where

\[ r(\hat{g}, x_j) = [(g^0 - \hat{g})' H_1(\hat{g}, x_j)(g^0 - \hat{g}), ..., (g^0 - \hat{g})' H_d(\hat{g}, x_j)(g^0 - \hat{g})]' \]

and \( F(\hat{g}, x_j) \) is the \( d \times r \) matrix of first partial derivatives of \( g(\hat{g}, x) \) evaluated at \( \hat{g} = \hat{g}, x = x_j \). Furthermore, since
\[ V^{-1}(\hat{\chi})[n^{-1} \sum_{j=1}^{n} F'(\hat{\varphi}, x_j) A^{-1}(\hat{\chi}, x_j) F(\hat{\varphi}, x_j)] = I, \]

It follows from (3.7.3) that

\[ \hat{g}^{(1,n)} - \hat{g}^0 = V^{-1}(\hat{\chi})[n^{-1} \sum_{j=1}^{n} F'(\hat{\varphi}, x_j) A^{-1}(\hat{\chi}, x_j) r(\hat{\varphi}, x_j)] + V^{-1}(\hat{\chi})[n^{-1} \sum_{j=1}^{n} F'(\hat{\varphi}, x_j) A^{-1}(\hat{\chi}, x_j) e_j] \quad (3.7.1.3) \]

Let \( y_\chi(\hat{\chi}) \) be the \( s \)-th element of

\[ n^{-1} \sum_{j=1}^{n} F'(\hat{\varphi}, x_j) A^{-1}(\hat{\chi}, x_j) r(\hat{\varphi}, x_j), \quad s = 1, 2, \ldots, r. \]

Then,

\[ y_\chi(\hat{\chi}) = n^{-1} \sum_{j=1}^{n} \sum_{t=1}^{d} \sum_{i=1}^{d} g_t(s)(\hat{\varphi}, x_j) b_{ti}(\hat{\chi}, x_j)(\hat{g}^0 - \hat{\varphi})' H_1(\hat{\varphi}, x_j)(\hat{g}^0 - \hat{\varphi}) \]

\[ = \frac{1}{2} \sum_{t=1}^{d} \sum_{i=1}^{d} \sum_{j=1}^{n} \sum_{q=1}^{r} g_t(s)(\hat{\varphi}, x_j) b_{ti}(\hat{\chi}, x_j) g^{(2k)}_i(\hat{\varphi}, x_j) \cdot (\hat{g}^0 - \hat{\varphi})_q \quad (3.7.1.4) \]

where \( b_{ti}(\hat{\chi}, x_j) \) is the \((t,i)\)-th element of \( A^{-1}(\hat{\chi}, x_j) \). Then, by Theorem 3.3.1, it follows that for each \( t,i = 1, 2, \ldots, d \) and each \( q = 1, 2, \ldots, r \)

\[ n^{-1} \sum_{j=1}^{n} g_t(s)(\hat{\varphi}, x_j) b_{ti}(\hat{\chi}, x_j) g^{(2k)}_i(\hat{\varphi}, x_j)' \]
converges uniformly in $\beta \in \Theta$, $\gamma \in \Phi$ to some constant, say $L_{t,i,\varepsilon,q}(\gamma)$, which by Theorem 2.5.1 is bounded. Then, for any $n_1 > 0$ there exists an integer $N_1$ such for $n > N_1$

$$|n^{-1} \sum_{j=1}^{n} g^{(s)}_c(\hat{\beta}, x_j)b_{t,1}(\gamma, x_j)g^{(\ell k)}_i(\hat{\beta}, x_j) - L_{t,i,\varepsilon,q}(\gamma)| < \eta_1$$

for all $\beta \in \Theta$, $\gamma \in \Phi$. (3.7.1.5)

As $\gamma - \gamma^0 = o_p(a_n)$ and $a_n \rightarrow 0$, given $n_2 > 0$, there exists an integer $N_2$ such that

$$\Pr[\hat{\beta} \in \mathbb{N}(\gamma^0), \gamma \in \mathbb{N}(\hat{\gamma}^0)] > 1 - n_2$$

whenever $n > N_2$. (3.7.1.6)

Then,

$$\Pr[|n^{-1} \sum_{j=1}^{n} g^{(s)}_c(\hat{\beta}, x_j)b_{t,1}(\gamma, x_j)g^{(\ell k)}_i(\hat{\beta}, x_j) - L_{t,i,\varepsilon,q}(\gamma)|^2 > \eta_1^2]$$

$$< \Pr[\hat{\beta} \notin \mathbb{N}(\gamma^0) \text{ or } \gamma \notin \mathbb{N}(\hat{\gamma}^0)] < n_2$$

and since $L_{t,i,\varepsilon,q}(\gamma^0)$ is a constant,

$$n^{-1} \sum_{j=1}^{n} g^{(s)}_c(\hat{\beta}, x_j)b_{t,1}(\gamma, x_j)g^{(\ell k)}_i(\hat{\beta}, x_j) = o_p(1)$$

Then, each term $y_s(\gamma)$ is $o_p(a_n^2)$, and
Note that each element of $V_n(\chi)$ is a continuous function of $\chi$, then

$$V_n(\chi) = V_n(\chi^0) + O_p(a_n).$$

Then, since $V_n(\chi)$ is nonsingular for sufficiently large $n$, $V_n^{-1}(\chi) = V_n^{-1}(\chi^0) + o_p(1)$. Consequently,

$$V_n^{-1}(\chi) = o_p(1).$$

Therefore, by (3.7.1.7) and (3.7.1.8)

$$V_n^{-1}(\chi)[n^{-1} \sum_{j=1}^{n} F'(\hat{\theta}, x_j)A^{-1}(\chi, x_j)x(\hat{\theta}, x_j)] = O_p(a_n^2).$$

For each $s = 1, 2, \ldots, r$, let $Z_s(\chi)$ be the $s$-th element of

$$n^{-1} \sum_{j=1}^{n} F'(\hat{\theta}, x_j)A^{-1}(\chi, x_j)e_j.$$

Then

$$Z_s(\chi) = \sum_{t=1}^{d} \sum_{i=1}^{d} n^{-1} \sum_{j=1}^{n} g(s)(\hat{\theta}, x_j)b_{ti}(\chi, x_j)e_{ij}$$

where $e_{ij} = P_{ij} - g_{ij}(\chi^0, x_j)$. Let $b_{ti}^*(\chi, x_j)$ denote the first partial derivative of $b_{ti}(\chi, x)$ with respect to $\gamma_t$, $t = 1, 2, \ldots, r+u$, evaluated at $\chi = \chi^*$, $x = x_j$. Similarly, defined the second partial derivative $b_{ti}^{*(lm)}(\chi, x_j)$, for each $t, m = 1, 2, \ldots, r+u$. Also, let $g_{t,s}(\chi, x_j)$ be the third partial...
derivative of \( g_t(\beta, x) \) evaluated at \( \beta = \hat{\beta}, \quad x = x_j \). Then, a Taylor's series expansion of

\[
\sum_{j=1}^{n} g_t(\beta_j, x_j) b_{t1}(\chi, x_j) e_{ij}
\]

around \( \chi^0 \) gives

\[
\sum_{j=1}^{n} g_t(\beta_j, x_j) b_{t1}(\chi, x_j) e_{ij} = \sum_{j=1}^{n} g_t(\beta^0_j, x_j) b_{t1}(\chi^0, x_j) e_{ij} + \sum_{j=1}^{n} g_t(\beta^0_j, x_j) b_{t1}(\chi^0, x_j) e_{ij}
\]

\[+ \sum_{l=1}^{u} \sum_{j=1}^{n} g_t(s)^{(s)}(\chi^0, x_j) e_{ij} (\gamma^0 - \hat{\gamma}_l) \]

\[+ \sum_{l=1}^{u} \sum_{m=1}^{n} \lambda_{t1}(\chi, x_j) e_{ij} (\gamma^0_m - \hat{\gamma}_m) \]

(3.7.1.11)

where

\[
\xi_{t1}^{(s)}(\chi^0, x_j) = g_t(\beta_j, x_j) b_{t1}(\chi^0, x_j) + g_t(\beta^0_j, x_j) b_{t1}(\chi^0, x_j),
\]

\[
\lambda_{t1}^{(s\beta)}(\chi, x_j) = g_t(\beta_j, x_j) b_{t1}(\chi, x_j) + g_t(\beta^0_j, x_j) b_{t1}(\chi^0, x_j)
\]

any partial derivative of \( g_t^{(s)}(\beta, x) \) for \( \ell = r+1, \ldots, r+u \) is zero and, \( \gamma = \text{vec}(\beta, \beta) \) is in the line segment joining \( \chi^0 \) and \( \hat{\chi} \). By Theorem 3.3.1, \( n^{-1} \sum_{j=1}^{n} [\xi_{t1}^{(s)}(\chi^0, x_j)]^2 b_{t1}(\chi^0, x_j) \) converges to some
constant that depends on $s$, $\varepsilon$, $t$, $i$ and $\chi^0$. Then, there exists a constant $M$ such that

$$n \mathbb{E}[\sum_{j=1}^{n-1} \xi_{ij}^2(x_j, x_j^*)e_{ij}] < M \text{ for all } n.$$ 

Then by Corollary 5.1.1.1 in Fuller (1976)

$$n^{-1} \sum_{j=1}^{n} \lambda_{ij}^{(s^m)}(\chi^0, x_j)e_{ij} = o_p(n^{-1/2}). \quad (3.7.1.12)$$

By Theorem 3.3.2, $n^{-1} \sum_{j=1}^{n} \lambda_{ij}^{(s^m)}(\chi, x_j)e_{ij}$ converges almost surely to zero, uniformly in $\chi \in \Theta \times \Phi$. Therefore,

$$n^{-1} \sum_{j=1}^{n} \lambda_{ij}^{(s^m)}(\chi^0, x_j)e_{ij} = o_p(1). \quad (3.7.1.13)$$

Then, by (3.7.1.10), (3.7.1.11), (3.7.1.12) and (3.7.1.13)

$$n^{-1} \sum_{j=1}^{n} F'(\tilde{g}, x_j)A^{-1}(\chi^0, x_j)e_j = n^{-1} \sum_{j=1}^{n} F'(\tilde{g}^0, x_j)A^{-1}(\chi^0, x_j)e_j + o_p(\max(n^{-1/2}a_n, s_n^2)). \quad (3.7.1.14)$$

Therefore, the theorem holds from (3.7.1.13), (3.7.1.9), (3.7.1.14), and from the fact that $\nu^{-1}(\nu) = \nu^{-1}(\nu^0) + o_p(1)$. \hfill \Box

If the order in probability of the error in the initial estimator $\hat{\chi}$ is not smaller than $n^{-1/2}$, then the order of error of the OSGN
estimator $\hat{\theta}^{(1,n)}$ is not larger than that in the original estimator.

The next result establishes the asymptotic normality of $\hat{\theta}^{(1,n)}$.

**Corollary 3.7.1.1.** Let $\hat{\theta}^{(1,n)}$ be the OSGN estimator defined in (3.7.3) and let $\hat{\chi} = \text{vec}(\hat{\theta}, \hat{\varphi})$ be a consistent estimator of $\chi^0 = \text{vec}(\theta^0, \varphi^0)$ such that $\hat{\chi} = \chi^0 + O_p(a_n)$ where $\lim n^{1/2} a_n = 0$.

Consider the model defined in (3.1.1)-(3.1.3) and the linear model defined in (3.4.1). Let Conditions (3.1)-(3.11) hold. Then

$$n^{1/2} (\hat{\theta}^{(1,n)} - \theta^0) \xrightarrow{d} N_p[0, \nu^{-1}(\chi^0)],$$

where $\nu^{-1}(\chi^0) = \lim n^{-1} \sum_{j=1}^{n} F'(\theta^0, x_j) A^{-1}(\chi^0, x_j) F(\theta^0, x_j) .

**Proof.** By Theorem 3.7.1

$$n^{1/2} (\hat{\theta}^{(1,n)} - \theta^0) = \nu^{-1}(\chi^0) n^{1/2} \sum_{j=1}^{n} F'(\theta^0, x_j) A^{-1}(\chi^0, x_j) e_j + O_p\left(\max(a_n, n^{1/2} a_n^2)\right) ,$$

and the result follows from Theorem 3.5.1.

The OSGN estimator $\hat{\theta}^{(1,n)}$ can be used in place of $\hat{\theta}$ in (3.7.3) and the Two-Step-Gauss-Newton estimator $\hat{\theta}^{(2,n)}$ can be computed. If this is done $k$-times, the $k$-Step-Gauss-Newton estimator $\hat{\theta}^{(k,n)}$ is well defined and, by Corollary 3.7.11, is asymptotically normal if

$$\lim n^{1/2} a_n = 0.$$
4. APPLICATIONS TO LOGIT MODELS

This chapter presents some applications and examples of the general model studied in Chapter III. In these particular applications the only link function defined in (3.1.3) considered is the logistic function. There are many ways of constructing logits with more than two categories. In this chapter, logits are constructed from the log-odds of being in the i-th category against the last one. Similar applications can be easily developed for other link functions.

4.1. A Logistic Multinomial Model

Consider the model (3.1.1)-(3.1.3) and assume that the conditional distribution of \( \pi_j = \pi_j^* \) given \( x_j \) is multinomial with sample size \( m_j \) and probability vector \( \pi_j^* \). Then

\[
E(\pi_j^*) = \pi_j^* , \quad j = 1, 2, \ldots, n \tag{4.1.1}
\]

and

\[
\text{Var}(\pi_j^*) = m_j^{-1} \Lambda(\pi_j^*) . \tag{4.1.2}
\]

One way of linking the \( \pi_j^* \) to the observed covariates \( x_j \) is via a generalized logistic function,

\[
\pi_j^* = f(a^0, x_j) , \quad j = 1, 2, \ldots, n \tag{4.1.3}
\]

where
\[ f(\beta^0, x_j) = \left[ \sum_{i=1}^{d+1} \exp(x_j \beta_{i0}) \right]^{-1} \left[ \exp(x_j \beta_{10}), \ldots, \exp(x_j \beta_{d0}) \right] . \quad (4.1.4) \]

This model is overparameterized, but a simple transformation leads to

\[ \eta^*_j = [g'(\beta^0, x_j), 1 - g'(\beta^0, x_j)]' \quad (4.1.5) \]

where

\[ g(\beta^0, x_j) = [1 + \sum_{i=1}^{d} \exp(x_j \beta_{i0})]^{-1} \left[ \exp(x_j \beta_{10}), \ldots, \exp(x_j \beta_{d0}) \right] . \quad (4.1.6) \]

\[ \beta^0 = (\beta_{10} - \beta_{d0})^{T}, \quad i=1, 2, \ldots, d \quad \text{and} \quad \beta^0 = \text{vec}(\beta_{10}, \ldots, \beta_{d0}) . \]

Henceforth, parameterization (4.1.5) of the logistic function will be adopted.

Let \( ^{\hat{\cdot}}(\eta) = ^{\hat{\beta}}_{\text{MLE}} \) denote the maximum likelihood estimator of \( \beta^0 \).

Consider the next two new conditions:

**Condition 4.1.**

\[ \lim_{n \to \infty} n^{-1} \sum_{j=1}^{d+1} \sum_{i=1}^{d} \pi_{ij}(\beta^0) \ln \left( \frac{\pi_{ij}(\beta)}{\pi_{ij}(\beta^0)} \right) < 0 \quad \text{for} \quad \beta \neq \beta^0, \beta \in \Theta . \]

**Condition 4.2.** There exists an \( \eta > 0 \) such that

\[ \lim_{n \to \infty} n^{-1} \sum_{j=1}^{d+1} \sup_{\beta \neq \beta^0, \beta \in \Theta} \sum_{i=1}^{d} \pi_{ij}(\beta^0) \ln \left( \frac{\pi_{ij}(\beta)}{\pi_{ij}(\beta^0)} \right) < 0 , \quad \text{for} \quad \|\beta\| > \eta . \]
where \( \| \hat{\beta} \| \) represents the Euclidean norm of \( \hat{\beta} \). Some asymptotic properties of \( \hat{\beta}_{\text{MLE}} \) are shown in the next theorem.

**Theorem 4.2.1.** Let the model (4.1.1)-(4.1.2) hold with link function (4.1.5). Under conditions (3.1), (3.3)-(3.7), (4.1)-(4.2),

1) \( \sqrt{n}(\hat{\beta}_{\text{MLE}}(n) - \beta^0) \xrightarrow{p} N_{dk}(0, \mathbb{V}^{-1}(\beta^0)) \) as \( n \to \infty \),

2) \( \hat{\beta}_{\text{MLE}}(n) = \left( n^{-1} \sum_{j=1}^{n} G'(\beta^0, x_j) \Delta^{-1}(x_j^*) G(\beta^0, x_j) \right)^{-1} \)

\[ \times n^{-1} \sum_{j=1}^{n} G'(\beta^0, x_j) \Delta^{-1}(x_j^*) [\bar{y}_j - x_j^*] + O_p(n^{-1/2}) \]

where \( G(\beta^0, x_j) \) represents the \((d+1) \times dk\) matrix of first partial derivatives of \( \pi_j = [g'(\beta, x), 1 - g'(\beta, x)] \) evaluated at \( \beta = \beta^0 \) and \( x = x_j \), and \( \Delta^{-1}(x_j^*) \) is any generalized inverse of \( \Delta(x_j^*) \).

**Proof of 1.** The true parameter vector \( \beta^0 = \text{vec}(\beta_1^0, \ldots, \beta_d^0) \) belongs, by Conditions (3.1) and (3.6), to the interior of a compact subset of \( \mathbb{R}^{dk} \) and the \( x_j \)'s are confined by Condition (3.5) to a compact set. Since the logistic function (4.1.5) is continuous with continuous first and second partial derivatives, then by Theorem (2.5.2) the elements of \( \pi_j^* \) are bounded away from zero and one, uniformly in \( j \), and the first and second partial derivatives of (4.1.5) are uniformly bounded for all \( j = 1, 2, \ldots \). In particular, the second partial derivatives are uniformly bounded in a neighborhood of \( \beta^0 \).
Let \( L_n(g^0) \) be the likelihood function of \( (T^n_1, \ldots, T^n_n) \) where the \( T^n_j = m_j \pi^n_j \) are independent multinomial random vectors with parameters \( [m_j, \pi^n_j] \). Since \( L_n(g^0) \) is proportional to

\[
\prod_{j=1}^{d+1} \prod_{i=1}^n \pi^n_j (g^0),
\]

the log likelihood function can be written as

\[
\ln(\pi^n_j) = (\ln \pi_{1j}, \ldots, \ln \pi_{d+1,j}).
\]

Let \( S_n(g^0) = \frac{\partial}{\partial g^0} \ln L_n(g^0) \) be the Fisher score vector and let \( I_n(g^0) = \text{Cov} S_n(g^0) \) be the Fisher information matrix. Then after some algebraic manipulations,

\[
\frac{\partial}{\partial g^0} l_j(g^0) = G'(g^0, x_j)(\text{Diag } \pi^n_j)^{-1} \pi^n_j.
\]

(4.2.1.1)

where \( G'(g^0, x_j) = \{[I_d, 0] \otimes x_j\} \Delta(\pi^n_j) \). But \( \pi^n_j \) is orthogonal to \( \Delta(\pi^n_j)(\text{Diag } \pi^n_j)^{-1} \), then

\[
\frac{\partial}{\partial g^0} l_j(g^0) = G'(g^0, x_j)(\text{Diag } \pi^n_j)^{-1}[\pi^n_j - \pi^n_j].
\]

(4.2.1.2)

Therefore,

\[
I_n(g^0) = \sum_{j=1}^n G'(g^0, x_j)(\text{Diag } \pi^n_j)^{-1}\Delta(\pi^n_j)(\text{Diag } \pi^n_j)^{-1}G(g^0, x_j).
\]

(4.2.1.3)

Since \( G(g^0, x_j) \) is in the column space of \( \Delta(\pi^n_j) \), it follows from (4.2.1.3) and Lemma 2.24 in Rao and Mitra (1971) that for any generalized inverse of \( \Delta(\pi^n_j) \)
In particular, this is true for the choice

$$
\Delta = \begin{pmatrix}
\Delta^{-1}(\pi_j) & 0 \\
0 & 0
\end{pmatrix}.
$$

Then

$$
I_n(\theta^0) = \sum_{j=1}^{n} F'(\theta^0, \pi_j)\Delta^{-1}(\pi_j)F(\theta^0, \pi_j).
$$

where $F_n(\theta^0, \pi_j)$ is the $d\times dk$ matrix of first partial derivatives of $\pi = g(\theta, \pi)$ with respect to $\theta$, evaluated at $\theta = \theta^0$ and $\pi = \pi_j$. Therefore, Condition (3.7) is exactly the same as

$$
\lim_{n} n^{-1} I_n(\theta^0) = W(\theta^0) \text{ with } W(\theta^0) \text{ being positive definite. Then, by Dale (1986),}
$$

$$
\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta^0) \xrightarrow{P} N_{dk}[0, W^{-1}(\theta^0)].
$$

Proof of ii). By Corollary 2.2.2 the elements of

$$
\Delta^{-1}(\pi_j) = (\text{Diag } \pi_j)^{-1} + \gamma J, \text{ where } \gamma = 1 - \pi_j'J, \text{ and } J = 1_{1'} \text{ are uniformly bounded for all } j. \text{ Then Condition (3.2) is satisfied.}
$$

Therefore, Theorem 3.5.1 implies that

$$
\sqrt{n}(\hat{\theta}^{(n)} - \theta^0) = \sqrt{n} W^{-1}(\theta^0) W_n(\theta^0) \xrightarrow{P} N_{dk}[0, W^{-1}(\theta^0)].
$$
where \( V_n(\theta^0) = n^{-1} \sum_{j=1}^{n} F'(\theta^0, x_j)\Delta^{-1}(\pi_j)F(\theta^0, x_j) \) and \( U_n(\theta^0) = n^{-1} \sum_{j=1}^{n} F'(\theta^0, x_j)\Delta^{-1}(\pi_j)[P_j - \pi_j] \). Since \( G(\theta^0, x_j) \) and \( [P_j - \pi_j] \) belong to the column space of \( \Delta(\pi_j) \), it follows from (4.2.1.2) that

\[
S_n(\theta^0) = \sum_{j=1}^{n} G'(\theta^0, x_j)\Delta^{-1}(\pi_j)[P_j - \pi_j]
\]

\[
= \sum_{j=1}^{n} F'(\theta^0, x_j)\Delta^{-1}(\pi_j)F(\theta^0, x_j)
\]

\[
= n U_n(\theta^0).
\] (4.2.1.8)

Therefore, by (4.2.1.4), (4.2.1.7) and (4.2.1.8)

\[
\sqrt{n}(\hat{\theta}^{(n)} - \theta^0) = \sqrt{n}[n^{-1} \sum_{j=1}^{n} G'(\theta^0, x_j)\Delta^{-1}(\pi_j)G(\theta^0, x_j)]^{-1}
\]

\[
\times (n^{-1} \sum_{j=1}^{n} G'(\theta^0, x_j)\Delta^{-1}(\pi_j)[P_j - \pi_j])
\]

\[
\xrightarrow{d} N_{dk}[0, V^{-1}(\theta^0)],
\] (4.2.1.9)

for any choice of \( \Delta^{-1}(\pi_j) \). The proof follows from Part 1) and (4.2.1.9).

A practical consequence of using the generalized inverse specified by (4.2.1.5) is that the OSGN estimator of equation (3.7.3) of Section 3.7 is
where \( \hat{\beta}_0 \) is a consistent estimator of \( \beta^0 \).

For the binomial case \((d = 1)\), the OSGN estimator of \( \beta^0 \) can be written as

\[
\hat{\beta}^{(1,n)} = \hat{\beta} + \left\{ \sum_{j=1}^{n} m_j \Delta(\hat{\pi}_j) \right\}^{-1} \left\{ \sum_{j=1}^{n} m_j (\hat{\pi}_j - \hat{\pi}_j) \right\} \cdot \hat{\beta},
\]

(4.1.7)

where \( \hat{\pi}_j = g(\hat{\beta}, x_j) \) and \( \hat{\beta} \) is a consistent estimator of \( \beta^0 \).

For the binomial case \((d = 1)\), the OSGN estimator of \( \beta^0 \) can be written as

\[
\hat{\beta}^{(1,n)} = (X'G)^{-1}X'G(\beta^0) + z,
\]

(4.1.8)

where \( X = \text{Block vec}(x_1', \ldots, x_n') \),  
\( G = \text{Diag} [m_1 \hat{\pi}_1 (1 - \hat{\pi}_1), \ldots, m_n \hat{\pi}_n (1 - \hat{\pi}_n)] \),  
\( z = \{[\hat{\pi}_1 (1 - \hat{\pi}_1)]^{-1} (P_1 - \hat{\pi}_1), \ldots, [\hat{\pi}_n (1 - \hat{\pi}_n)]^{-1} (P_n - \hat{\pi}_n)\}' \),  
and \( \hat{\pi}_j = g(\hat{\beta}, x_j) \) for \( j = 1, 2, \ldots, n \). Nelder and Wedderburn (1972) have shown that under the binomial assumption, the likelihood equations can be solved by iteratively applying formula (4.1.7).

Haberman (1974) showed that under regularity conditions a modified Newton-Raphson converges to the maximum likelihood estimator \( \hat{\beta}_{\text{MLE}} \) for the multinomial case \((d > 2)\). His proof does not depend on the existence of any consistent estimator of \( \beta^0 \) which allows the iterative algorithm to be initialized at \( \hat{\beta} = 0 \). This initial solution is convenient when the sample size \( m_j \) for each multinomial \( T_j \) is small or one. If \( m_j \) are large an initial estimator \( \hat{\beta} \) can be obtained using the weighted least squared method described in Grizzle, Starmer and Koch (1969). This estimator is computed by regressing the empirical logist against the covariates, i.e.,

\[
\hat{\beta} = H_n^{-1}W_n
\]

(4.1.9)
where $H_n = \sum_{j=1}^{n} P_j \Delta^{-1}(P_j) R_j$, $W_n = \sum_{j=1}^{n} m_j P_j \Delta^{-1}(P_j) y_j$, $R_j$ is the matrix of partial derivatives of $[\ln(\pi_1/\pi_{d+1}), \ldots, \ln(\pi_d/\pi_{d+1})]'$ with respect to $\hat{\beta}_j$ evaluated at $\hat{\theta} = P_j$, and $y_j = [\ln(P_{1j}/P_{d+1,j}), \ldots, \ln(P_{dj}/P_{d+1,j})]'$. If for some $j$ a category has zero observed frequency then the empirical logit is not defined. In such a case, the usual prescription is to increase the empty cell and the sample size $m_j$ by a small quantity like 0.5. Using such an adjustment

$$H_n = \sum_{j=1}^{n} m_j \Delta(P_j) x_j x_j'$$

and

$$W_n = \sum_{j=1}^{n} m_j \Delta(P_j) y_j x_j$$

Jennrich and Moore (1975) and Bradley (1973) have pointed out that the solution of the maximum likelihood equations is equivalent to an iteratively reweighted generalized least squares procedure with weights changing on each iteration. Jennrich and Moore (1975) proved that the common Gauss-Newton algorithm for nonlinear least squares becomes the Fisher scoring algorithm for maximum likelihood estimation for the exponential family. In the multinomial logistic case, it is also the Newton-Raphson algorithm. Because of this equivalence of those algorithms and because a modified Newton-Raphson procedure always converges, it will be adopted henceforth, a modified Gauss-Newton algorithm for computing $\hat{\beta}_{\text{MLE}}$ under the logistic multinomial model.
The algorithm can be described as follows.

Step 0: Obtain an initial solution \( \hat{\beta} \) either by making \( \hat{\beta} = 0 \) or by computing \( \hat{\beta} \) from (4.1.9). Evaluate the log likelihood 
\( \ell(\hat{\beta}) = \ln L_n(\hat{\beta}) \). If the initial solution is the zero vector, the log likelihood is \(- n \ln(d + 1)\) where \( n \) is the number of multinomials and \( (d + 1) \) the number of categories.

Step 1: Define
\[
\hat{\gamma}_n(\hat{\beta}) = \left[ \sum_{j=1}^{n} m_j \Delta(\hat{\pi}_j) \right]^{-1} \left[ \sum_{j=1}^{n} m_j (\hat{P}_j - \hat{\pi}_j) \right] \hat{x}_j
\]
where \( \hat{\pi}_j = g(\hat{\beta}, x_j) \). Then compute
\[
\hat{\beta}^{(i)} = \hat{\beta} + (0.5)^i \hat{\gamma}_n(\hat{\beta})
\]
and
\[
\ell(\hat{\beta}^{(i)}) \quad i=0, 1, 2, \ldots
\]
until an integer \( i^* \) is found such that \( \ell(\hat{\beta}^{(i^*)}) > \ell(\hat{\beta}) \).

Replace \( \hat{\beta} \) with \( \hat{\beta}^{(i^*)} \) and \( \ell(\hat{\beta}) \) with \( \ell(\hat{\beta}^{(i^*)}) \). Then repeat Step 1. Gallant (1987) has suggested if the condition
\[
\frac{\ell(\hat{\beta}^{(i^*)}) - \ell(\hat{\beta})}{|\ell(\hat{\beta}^{(i^*)})| + 10^{-3}} < 10^{-5}
\]
(4.1.14)
is satisfied then the algorithm converges and must be stopped.

4.2. A Scaled Multinomial Variance Model

Let the model (3.1.1)-(3.1.3) hold and assume that

\[ E(\mathbf{P}_j) = \pi_j \]
\[ = g(x^0, x_j) \]  \hspace{1cm}  (4.2.1)

and

\[ \text{Var}(\mathbf{P}_j) = m_j^{-1} \phi^0 \Lambda(\pi_j) \]  \hspace{1cm}  (4.2.2)

where \( g(x^0, x_j) \) is the generalized logistic function defined in (4.1.6), and \( \phi^0 \) is a positive constant. The logistic multinomial model described in Section 4.1 is obtained when \( \phi^0 = 1 \).

For any given \( \pi_j \), there exists a matrix \( E_j \) such that

\[ \Lambda(\pi_j) = E_j \Lambda_j E_j' \]  \hspace{1cm}  (4.2.3)

where \( E_j E_j' = I = E_j' E_j \) and \( \Lambda_j = \text{Diag}(\lambda_{1j}, \ldots, \lambda_{dj}) \), \( \lambda_{ij} > 0 \) for \( i = 1, 2, \ldots, d \). Then

\[ \Lambda(\pi_j) = \Delta^{1/2}(\pi_j) \Delta^{1/2}(\pi_j) \]  \hspace{1cm}  (4.2.4)

where \( \Delta^{1/2}(\pi_j) = E_j^{1/2} \), \( \Delta^{1/2}(\pi_j) = \text{Diag}(\lambda_{ij}^{1/2}, \ldots, \lambda_{dj}^{1/2}) \). Define
\[ z_j = m_j^{1/2} \Delta^{-1/2} (\pi_j - \bar{\pi}_j) \tag{4.2.5} \]

where \( \Delta^{-1/2} (\pi_j) = E_j \Delta^{-1/2} E_j' \), \( \Delta^{-1/2} = \text{Diag}(\lambda_{1j}^{-1/2}, \ldots, \lambda_{dj}^{-1/2}) \). Then, under the model assumptions (4.2.1)-(4.2.2)

\[ E(z_j) = 0 \tag{4.2.6} \]

and

\[ \text{Var}(z_j) = \phi^0 I . \tag{4.2.7} \]

So, a consistent estimator of \( \phi^0 \) is given by

\[ \hat{\phi} = [(n - k)d]^{-1} \text{Trace} \left( \sum_{j=1}^{n} z_j z_j' \right) . \tag{4.2.8} \]

In practice the \( \pi_j \)'s are unknown and a consistent estimator of \( \phi^0 \), \( \hat{\phi} \), is available, so the \( z_j \)'s are based on \( \hat{\pi}_j \)'s.

Now observe that

\[ \hat{\phi} = [(n - k)d]^{-1} \sum_{j=1}^{n} \text{trace}(z_j z_j') \]

\[ = [(n - k)d]^{-1} \sum_{j=1}^{n} m_j \text{trace}(\Delta^{-1/2} (\hat{\pi}_j - \bar{\pi}_j)[P_j - \hat{\pi}_j][P_j - \hat{\pi}_j]' \Delta^{-1/2} (\hat{\pi}_j)) \]

\[ = [(n - k)d]^{-1} \sum_{j=1}^{n} m_j [P_j - \hat{\pi}_j]' \Delta^{-1} (\hat{\pi}_j) [P_j - \hat{\pi}_j] \]
So, two conclusions can be drawn from this last expression. The first one is that \( \hat{\phi} \) is invariant to the choice of any generalized inverse of \( \Delta(\pi_j) \) and the second one is that \( \hat{\phi} \) can be computed as 

\[
(n - k)d^{-1}nQ(\hat{\phi}), \quad \text{where} \quad Q(\hat{\phi}) = \sum_{j=1}^{n-1} m_j [P_j - \pi_j]^{(n - k)d} \{\Delta(\pi_j)\}^{-1}[P_j - \pi_j],
\]

which requires no spectral decomposition.

Let \( \hat{\theta} \) and \( \hat{\phi} \) be consistent estimators of \( \theta^0 \) and \( \phi^0 \) respectively. Then the OSGN estimator of \( \theta^0 \) is defined as

\[
\hat{\theta}^{(1, n)} = \hat{\theta} + \left[ \hat{\phi} V_n(\hat{\theta}) \right]^{-1} \hat{\theta} U_n(\hat{\theta})
\]

\[
= \hat{\theta} + \left[ V_n(\hat{\theta}) \right]^{-1} U_n(\hat{\theta})
\]

(4.2.10)

where

\[
V_n(\hat{\theta}) = n^{-1} \sum_{j=1}^{n} \left( \theta_j - \hat{\theta}_j \right) \Delta^{-1}(\pi_j) \{g(\hat{\theta}, x_j)\} \Delta^{-1}(\pi_j) \{g(\hat{\theta}, x_j)\}
\]

and

\[
U_n(\hat{\theta}) = n^{-1} \sum_{j=1}^{n} \left( \theta_j - \hat{\theta}_j \right) \Delta^{-1}(\pi_j) \{g(\hat{\theta}, x_j)\} \Delta^{-1}(\pi_j) \{g(\hat{\theta}, x_j)\}.
\]

This expression does not depend on the value of \( \hat{\phi} \) and coincides completely with the OSGN estimator based on the logistic multinomial
model defined in (4.1.7). Therefore, if one uses the modified Gauss-Newton algorithm described in (4.1.12)-(4.1.14), the iterative procedure should converge to the maximum likelihood estimator for the logistic multinomial model. If one assumes model (4.1.1)-(4.1.6) when in fact model (4.2.1)-(4.2.2) holds, the resulting estimator will be called a pseudo maximum likelihood estimator and denoted by $\hat{\beta}_{\text{PSEUDO}}$. By Theorem (4.1.1), under model (4.2.1)-(4.2.2), $\sqrt{n}(\hat{\beta}_{\text{PSEUDO}} - \beta^0)$ converges in law to a multivariate normal $d_k$-dimensional random vector with mean 0 and covariance matrix $[\phi^0 \Psi(\beta^0)]^{-1}$, where

$$\Psi(\beta^0) = \lim_{n \to \infty} \Psi_n(\beta^0).$$

If one uses $\hat{\beta}_{\text{PSEUDO}}$, assuming the logistic multinomial model when in fact the (4.2.1)-(4.2.2) holds, then the chi-square test for testing $H_0: A\beta^0 = \gamma^0$, where $A$ is any $r \times d_k$ full row rank matrix of known coefficients and $\gamma^0$ is a hypothesized column vector of dimension $r$, must be scaled by the factor $\phi^0$. If for instance, $\phi^0$ is greater than one and is completely ignored, the chi-square test based on the logistic multinomial model will tend to have a type I error level greater than the nominal level. Similarly, the covariance matrix of $\hat{\beta}_{\text{PSEUDO}}$ will tend to be underestimated by the factor $(\phi^0)^{-1}$.

A test to check the model assumption can be computed from $\hat{\beta}_{\text{PSEUDO}}$. If $\phi^0$ was known, the estimator

$$[(n - k)d] \phi(\phi^0)^{-1} = \sum_{j=1}^{n} m_j [P_j - g(\hat{\beta}_{\text{PSEUDO}}, x_j)]' \Delta^{-1} [g(\hat{\beta}_{\text{PSEUDO}}, x_j)]$$

$$\times [P_j - g(\hat{\beta}_{\text{PSEUDO}}, x_j)]$$
has approximately chi-square distribution with \((n - k)d\) degrees of freedom. So, under the logistic multinomial model, \(\phi\) should be close to one.

4.3. A Generalized Multinomial Variance Model

Let the model (3.1.1)-(3.1.3) hold with

\[
E(P_j) = \pi_j
\]

\[
= g(\xi_j, x_j), \tag{4.3.1}
\]

where \(g(\xi_j, x_j)\) is the generalized logistic function defined in (4.1.6),

\[
\text{Var}(P_j) = m_j^{-1} \Delta^{1/2} (\pi_j) \xi_j^{0.5} (\pi_j), \tag{4.3.2}
\]

and \(\xi_j^0\) is a \(d \times d\) unknown positive definite covariance matrix. The matrix \(\xi_j^0\) plays the role of the extra parameters described in Section (3.7). The simple situation, where \(\xi_j^0\) is a multiple of the identity matrix, has been studied in Section (4.2). Two new cases are presented here: where \(\xi_j^0\) is an unspecified \(d \times d\) matrix, and where \(\xi_j^0\) is a diagonal covariance matrix.

Under model assumptions (4.3.1)-(4.3.2) with \(\xi_j^0\) known, an unbiased and consistent estimator of \(\xi_j^0\) is
\[ \hat{\phi} = (n - dk)^{-1} \sum_{j=1}^{n} z_j z_j' \] (4.3.4)

where the \( z_j \) are based on \( P_j \) and \( \tilde{x}_j = g(\hat{\phi}, x_j) \). One way of evaluating \( \hat{\phi} \) initially substitutes \( \hat{\phi}_{\text{PSEUDO}} \) for \( \hat{\phi} \). If the maximum likelihood estimator of model (4.2.1)-(4.2.6) is computed iteratively using (4.1.12)-(4.1.14), when, in fact model (4.3.1)-(4.3.2) holds, then by Theorem (4.2.1) it follows that

\[ \sqrt{n}(\hat{\phi}_{\text{PSEUDO}} - \phi^0) \overset{d}{\longrightarrow} N_{dk}(0, H^{-1}(\phi^0)) \] (4.3.5)

where \( H^{-1}(\phi^0) = \lim[I_n(\phi^0)]^{-1} B_n(\phi^0, \phi^0) [I_n(\phi^0)]^{-1} \) almost surely,

\( I_n(\phi^0) \) is defined as in (4.2.1.4), and

\[ B_n(\phi^0, \phi^0) = \sum_{j=1}^{n} m_j \Delta^{1/2} (\tilde{x}_j)(\phi^0)^{-1} \Delta^{1/2} (\tilde{x}_j) \]

Consider the OSGN estimator of Section (3.7) with \( \chi^0 = \text{vec}[\phi^0, \text{vech}(\phi^0)] \). A way of computing this estimator can be described as follows.

Initially compute \( \hat{\phi}_{\text{PSEUDO}} \) and then \( \hat{\phi} = (n - dk) \sum_{j=1}^{n} z_j z_j' \) where
\[ \tilde{z}_j = m_j^{1/2} \tilde{\Delta}^{-1/2} (\tilde{\pi}_j) [\tilde{P}_j - \tilde{\pi}_j], \quad \tilde{\pi}_j = \mathbf{g}(\hat{\beta}_{\text{PSEUDO}}, \tilde{x}_j). \]

Based on \( \hat{\chi} = \text{Vec}(\hat{\beta}_{\text{PSEUDO}}, \text{vech}(\hat{\Phi})) \) compute

\[ \hat{\epsilon}(1,n) = \hat{\beta}_{\text{PSEUDO}} + [\nabla_n(\hat{\chi})]^{-1} u_n(\hat{\chi}) \quad (4.3.6) \]

where

\[ \nabla_n(\hat{\chi}) = n^{-1} \sum_{j=1}^{n} m_j \mathbf{F}'(\hat{\beta}_{\text{PSEUDO}}, \tilde{x}_j) \Delta^{-1/2} (\tilde{\pi}_j)(\hat{\Phi})^{-1} \Delta^{-1/2} (\tilde{\pi}_j) \mathbf{F}(\hat{\beta}_{\text{PSEUDO}}, \tilde{x}_j), \]

\[ u_n(\hat{\chi}) = n^{-1} \sum_{j=1}^{n} m_j \mathbf{F}'(\hat{\beta}_{\text{PSEUDO}}, \tilde{x}_j) \Delta^{-1/2} (\tilde{\pi}_j)(\hat{\Phi})^{-1} \Delta^{-1/2} (\tilde{\pi}_j) [\tilde{P}_j - \tilde{\pi}_j], \]

and \( \tilde{z}_j = \mathbf{g}(\hat{\beta}_{\text{PSEUDO}}, \tilde{x}_j) \).

The case for which \( \hat{\Phi} \) is a diagonal matrix, \( \hat{\Phi} = \text{Diag}(\hat{\Phi}) \), can be viewed as follows: For a given \( \tilde{\pi} \), consider the spectral decomposition of \( \Delta(\tilde{\pi}) \). Then

\[ \Delta(\tilde{\pi}) = \mathbf{E} \Lambda \mathbf{E}' \]

\[ = \sum_{i=1}^{d} \lambda_i \tilde{e}_i \tilde{e}_i' \quad (4.3.7) \]

where \( \Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_d) \), \( \mathbf{E} = (\tilde{e}_1, \ldots, \tilde{e}_d) \), \( \lambda_i \) is the \( i \)-th eigenvalue of \( \Delta(\tilde{\pi}) \), \( \tilde{e}_i \) is the eigenvector associated with \( \lambda_i \), and \( \mathbf{E}'\mathbf{E} = \mathbf{I} \). Then
\[ \Delta(\pi) = \Delta^{1/2}(\pi) \Delta^{1/2}(\pi) \]  \hspace{1cm} (4.3.8)

where \( \Delta^{1/2}(\pi) = \text{Ei}^{1/2} \) and \( \Delta^{1/2} = \text{Diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2}) \). If 
\[ \phi^0 = \text{Diag}(\phi^0), \]
then
\[ \Delta^{1/2}(\pi) \text{Diag}(\phi^0) \Delta^{1/2}(\pi) = \sum_{i=1}^{d} \phi^0_i \lambda_i e_i e_i'. \]  \hspace{1cm} (4.3.9)

Therefore, if each \( \phi^0_i \) is greater than one (or each is smaller than one) then the diagonal elements in (4.3.2) are greater (or smaller) than the diagonal elements of \( \Delta(\pi) \) implying extra (or less) variability than that which can be attributed to the multinomial distribution.

When \( \phi^0 \) is a diagonal matrix, the algorithm described in (4.3.6) can be implemented replacing \( \hat{\phi} \) in (4.3.4) with 
\[ (n - dk)^{-1} \sum_{j=1}^{n} \text{Diag}(z_j z_j'). \]

For large \( m_j \), the \( z_j \) are nearly identically distributed normal \((0, \phi^0)\) random vectors under the model assumption. Then approximate tests for \( H_0: \phi^0 = I \) and \( H_0: \phi^0 = \phi^0 I \) can be computed using the formulas given in Sections 10.7 and 10.8 in Anderson (1984).

4.4. An Extended Scaled Multinomial Variance Model

Assume that model (3.1.1)-(3.1.3) holds with

\[ E(z_j) = z_j \]
\[ = g(\phi^0, x_j) \]  \hspace{1cm} (4.4.1)
\[ \text{Var}(\mathbf{\bar{p}}_j) = c_j m_j^{-1} \Delta(\bar{\pi}_j) \quad (4.4.2) \]

where \( g(\theta, x) \) is the generalized logistic function defined in (4.1.6) and \( c_j \) a positive constant which may vary from macro unit to macro unit. Models of this form where \( c_j = c \) have been presented by Brier (1980) and McCullagh and Nelder (1983). Assuming that the \( \pi_j \) follow a Dirichlet distribution with mean vector \( \bar{\pi} \), it can be shown that the covariance matrix of the unconditional vector of observed proportions is a multiple of the covariance matrix of the multinomial distribution with vector of proportions \( \bar{\pi} \). More recently, Efron (1986), formulated models for analyzing binomial and Poisson data by using double exponential families. He assumed a logistic-like model for the unknown multipliers with an upper bound specified by the analyst.

Here it will be assumed that

\[ c_j = M f(\theta^0, \mathbf{x}_j) , \quad (4.4.3) \]

where \( M \) is an unknown constant, \( \theta^0 \) a set of \( s \)-dimensional vector of unknown parameters and \( f(\cdot) \) is a uniformly continuous cumulative distribution function. The constant \( M \), confined to a compact subset of \( \mathbb{R}^+ \), is to be estimated. In this presentation, the only function \( f(\cdot) \) that is considered is the logistic cumulative distribution function. It may be desirable to let the strictly increasing function
$f(\cdot)$ go up and down as the $x_j$'s vary. This can be accomplished by evaluating $f(\cdot)$ at points corresponding to a polynomial function of the $x_j$'s of sufficiently high order. To simplify the present discussion it is further assumed that the $x_j$'s are such that there is no need to consider more than linear function of the $x_j$'s. Then, $\xi^0$ is a $k$-dimensional vector, which as in the preceding models, will be confined to a proper compact set.

Let $\xi^0 = (\xi^0, \mu)'$, and let

$$
\xi_j = n^{1/2} \Delta^{-1/2} (\mu_j)(P_j - \mu_j).
$$

Then, under the model assumption

$$
E(\xi_j) = 0
$$

and

$$
\text{Var}(\xi_j) = Mf(\xi^0, x_j)I.
$$

A consistent estimator of $\xi^0$ can be obtained by choosing the value of $\xi^0$ that minimizes

$$
S_n(\xi^0) = \sum_{j=1}^{n} [d^{-1} \xi_j' z_j - Mf(\xi^0, x_j)]^2.
$$

This is a nonlinear regression problem with $\xi^0$ being the parameter of interest. Using the same kind of arguments as those in
Chapter 3, it can be proved that if \( \hat{\theta} \) is a consistent estimator of \( \theta^0 \) such that \( \hat{\theta} = \theta^0 + O_p(a_n) \) with \( \lim n a_n = 0 \), then the OSGN estimator of \( \theta^0 \),

\[
\hat{\theta}^{(1, n)} = \hat{\theta} + [V_n(\hat{\theta})]^{-1} n^{-1} \sum_{j=1}^{n} d'(\hat{\theta}, x_j)(d^{-1} z_j^T x_j - Mf(\hat{\theta}, x_j)) \tag{4.4.7}
\]

is \( N_{k+1}(0, [V(\theta^0)]^{-1}) + O_p[\max(n^{-1/2} a_n, a_n^2)] \). Here, \( d(\hat{\theta}, x_j) \) is a \((k + 1)\) dimensional row vector of first partial derivatives of \( Mf(\hat{\theta}, x) \) evaluated at \( \hat{\theta} = \hat{\theta} \) and \( x = x_j \),

\[
V_n(\hat{\theta}) = n^{-1} \sum_{j=1}^{n} d'(\hat{\theta}, x_j)d(\hat{\theta}, x_j)
\]

and

\[
V(\theta^0) = \lim n^{-1} \sum_{j=1}^{n} d'(\theta^0, x_j)d(\theta^0, x_j) \text{ almost surely.}
\]

A consistent estimator of \( \theta^0 \) can be calculated as follows. Under the logistic multinomial model compute the pseudo maximum likelihood estimator \( \hat{\theta}_{\text{pseudo}} \) according to algorithm (4.1.12)-(4.1.14). Fix \( \hat{\theta}_{\text{pseudo}} \) and define \( \delta_n(\hat{\theta}) \) as

\[
\delta_n(\hat{\theta}) = [V_n(\hat{\theta})]^{-1} n^{-1} \sum_{j=1}^{n} d'(\hat{\theta}, x_j)(d^{-1} z_j^T x_j - 1) \tag{4.4.8}
\]

where

\[
V_n(\hat{\theta}) = n^{-1} \sum_{j=1}^{n} d'(\hat{\theta}, x_j)d(\hat{\theta}, x_j),
\]
\[ z_j = m_j \frac{1}{2} \Delta - \frac{1}{2} (\hat{y}_j)(p_j - \hat{z}_j), \hat{z}_j = \mathcal{E}(\hat{\theta}_{\text{PSEUDO}}; x_j) \] and \( \hat{\theta} \) is an initial solution of \( \hat{\theta}^0 \). A possible value of \( \hat{\theta} \) could be \( (0', 2) \).

With these values, \( c_j \) in (4.4.3) is always one and the first two moments of model (4.4.1)-(4.4.3) match with the first two moments of the logistic multinomial model of Section 4.1. Then compute

\[ \hat{\theta}^{(i)} = \hat{\theta} + (0.5)^i \hat{\delta}(\hat{\theta}), \quad i=1, 2, \ldots, \quad (4.4.9) \]

until an integer \( i^* \) is found such that

\[ S_n(\hat{\theta}) < S_n(\hat{\theta}^{(i^*)}) \quad (4.4.10) \]

where \( S_n(\hat{\theta}) \) is defined in (4.4.6).

Then replace \( \hat{\theta} \) with \( \hat{\theta}^{(i^*)} \) and repeat (4.4.9)-(4.4.10) until

\[ \frac{S_n(\hat{\theta}) - S_n(\hat{\theta}^{(i^*)})}{S_n(\hat{\theta}^{(i^*)}) + 10^{-3}} < 10^{-5} \quad (4.4.11) \]

Let \( \hat{\theta} \) be the estimator that is obtained when the modified Gauss-Newton method (4.4.8)-(4.4.11) is used. Using the initial estimates \( \hat{\theta}_{\text{PSEUDO}} \) and \( \hat{\theta} \), by Theorem 3.7.1, the OSGN estimator of \( \theta^0 \) is
\[ \hat{\beta}^{(1,n)} = \hat{\beta}_{\text{PSEUDO}} + \left[ \nabla_n(\hat{\beta}_{\text{PSEUDO}}, \hat{\phi}) \right]^{-1} \sum_{j=1}^{n} m_j F'(\hat{\beta}_{\text{PSEUDO}}, \mathbf{x}_j) \times [Mf(\hat{\phi}, \mathbf{x}_j)]^{-1} (\hat{\pi}_j - \pi_j) \] (4.4.12)

where

\[ \nabla_n(\hat{\beta}_{\text{PSEUDO}}, \hat{\phi}) = n^{-1} \sum_{j=1}^{n} F'(\hat{\beta}_{\text{PSEUDO}}, \mathbf{x}_j) m_j [Mf(\hat{\phi}, \mathbf{x}_j)]^{-1} (\hat{\pi}_j - \pi_j) F(\hat{\beta}_{\text{PSEUDO}}, \mathbf{x}_j) \]

and \( \hat{\pi}_j = g(\hat{\beta}_{\text{PSEUDO}}, \mathbf{x}_j) \).

A consideration in applying this model is that the distinction between misspecification of the model and random variation may become obscured. If \( f(\cdot) \) is allowed to be a sufficiently complicated function of the \( x_j \)'s, then the \( c_j \)'s may assume large enough values to prevent detection deficiencies in the specification of the model. The consequences may be that a model with a poorly specified mean component may be misjudged to fit well. When this occurs, the standard deviations of some parameter estimates will be very large, producing wide confidence intervals, and the power of test of hypotheses involving those parameters will be greatly reduced. If this situation is suspected, a new model could be tried using a smoother form of \( f(\cdot) \).

A generalization of model (4.4.1)-(4.4.3) is the following

\[ E(\mathbf{p}_j) = \pi_j \]

\[ = g(g^0, \mathbf{x}_j) \] (4.4.13)
and

\[ \text{Var}(\mathbf{P}_j) = m_j^{-1} \Delta^{1/2} (\pi_j) \text{Diag}(c_{1j}, \ldots, c_{dj}) \Delta^{1/2} (\pi_j), \]  

(4.4.14)

where \( c_{ij} = M_i f(\hat{\theta}_i, x_j) \) for \( i = 1, \ldots, d \). The algorithm (4.4.8)-(4.4.11) can be used for obtaining \( \hat{\varphi}_i = (\hat{\theta}_i, M_i), i=1, 2, \ldots, d \). Then the OSGN estimator of \( \varphi^0 \) is

\[ \hat{\varphi}^{(1,n)} = \hat{\varphi}_{\text{PSEUDO}} + \left[ \text{V}_n(\hat{\varphi}_{\text{PSEUDO}}, \hat{\varphi}_1, \ldots, \hat{\varphi}_d) \right]^{-1} \left[ \text{U}_n(\hat{\varphi}_{\text{PSEUDO}}, \hat{\varphi}_1, \ldots, \hat{\varphi}_d) \right], \]

(4.4.15)

where

\[ \text{V}_n(\hat{\varphi}_{\text{PSEUDO}}, \hat{\varphi}_1, \ldots, \hat{\varphi}_d) \]

\[ = n^{-1} \sum_{j=1}^{n} m_j F(\hat{\varphi}_{\text{PSEUDO}}, x_j) \Delta^{1/2} (\pi_j) [\text{Diag}(M_i f(\hat{\theta}_i, x_j), \ldots, \hat{M}_d f(\hat{\theta}_d, x_j))]^{-1} \Delta^{1/2} (\pi_j) F(\hat{\varphi}_{\text{PSEUDO}}, x_j), \]  

and

\[ \text{U}_n(\hat{\varphi}_{\text{PSEUDO}}, \hat{\varphi}_1, \ldots, \hat{\varphi}_d) \]

\[ = n^{-1} \sum_{j=1}^{n} m_j F(\hat{\varphi}_{\text{PSEUDO}}, x_j) \Delta^{1/2} (\pi_j) [\text{Diag}(M_i f(\hat{\theta}_i, x_j), \ldots, \hat{M}_d f(\hat{\theta}_d, x_j))]^{-1} \Delta^{1/2} (\pi_j) [F(\hat{\varphi}_{\text{PSEUDO}}, x_j) - \pi_j]. \]
4.5. A Model with a Finite Number of Different Covariance Matrices

Consider the case where the \( x_j \)'s take a finite number of different values \( x_{(1)}, x_{(2)}, \ldots, x_{(T)} \). For each \( t = 1, 2, \ldots, T \), let \( n_t \) represent the number of times \( x_{(t)} \) appears in the sequence \( x_1, x_2, \ldots, x_n \). A situation like this can arise when a designed experiment is conducted with \( x_{(1)}, x_{(2)}, \ldots, x_{(T)} \) representing the different combinations of the levels of the factors and, \( n_1, n_2, \ldots, n_T \) the number of times those different combinations are replicated.

Observe that Condition (3.3) assures that \( \lim n^{-1} \sum_{t=1}^{T} n_t \) exists, where \( n = \sum_{t=1}^{T} n_t \). It will be assumed that each limit is positive.

For each different \( x_{(t)} \), \( t = 1, 2, \ldots, T \), consider the vectors of proportions \( P_{t1}, P_{t2}, \ldots, P_{tn_t} \) and assume that

\[
E(P_{t\lambda}) = g(\beta^0, x_{(t)})
\]

\[
= \pi_{(t)} \text{ for } \lambda = 1, 2, \ldots, n_t , \quad (4.5.1)
\]

and

\[
\text{Var}(P_{t\lambda}) = m_{t\lambda} \Lambda_{(t)}
\]

\[
= \Lambda_{t\lambda} , \quad (4.5.2)
\]
where \( g(\hat{\beta}, x) \) is the generalized logistic function defined in (4.2.4) and \( \Sigma(t) \) is unspecified.

If \( \hat{\beta} \) is a consistent estimator of \( \beta^0 \) then a consistent estimator of \( \Sigma(t) \) is given by

\[
\hat{\Sigma}(t) = (n_t - dk)^{-1} \sum_{t=1}^{n_t} m_{t\ell} (\hat{\pi}_{t\ell} - \hat{\pi}(t)) (\hat{\pi}_{t\ell} - \hat{\pi}(t))',
\]

(4.5.3)

where \( \hat{\pi}(t) = g(\hat{\beta}, x(t)) \). If the underlying model is the generalized logistic multinomial, then \( \hat{\Sigma}(t) \) converges in probability to \( \Lambda(\pi(t)) \) as \( n_t \) increases. But, if it is suspected that the multinomial model does not hold, then \( \hat{\Sigma}(t) \) will converge in probability to the true covariance matrix \( \sigma(t) \). An attractive feature of this model is that the exact form of \( \Sigma(t) \) need not be specified. On the other hand, a disadvantage is that \( \Sigma(t) \) may not be well estimated if the corresponding value of \( n_t \) is small.

Let \( \hat{\beta}_\text{PSEUDO} \) and \( \hat{\Sigma}(t), \ t = 1, 2, ..., T \) be the initial estimators for this model. Then, by Theorem 3.7.1 the OSGN estimator of \( \beta^0 \) is

\[
\hat{\beta}^{(1,n)} = \hat{\beta} + [V_n(\hat{\beta})]^{-1} \sum_{t=1}^{T} m_{t\ell} F'(\hat{\beta}, x(t))^{-1} \Lambda(t)[\sum_{\ell=1}^{T} m_{t\ell} \pi_{t\ell} - M_{t} g(\hat{\beta}, x(t))]
\]

(4.5.4)

where \( V_n(\hat{\beta}) = n^{-1} \sum_{t=1}^{T} m_{t\ell} F'(\hat{\beta}, x(t))^{-1} \Lambda(t) F(\hat{\beta}, x(t)) \) and \( M_{t} = \sum_{\ell=1}^{T} m_{t\ell} \).
4.6. A Multinomial Variance Model under Complex Sampling Designs

In this section an estimation procedure is presented for obtaining consistent estimators of the parameter vector of a generalized logistic model and its asymptotic covariance matrix when a complex sampling design is employed. Since this estimation procedure has been incorporated into PC CARP [see Fuller et al. (1987)], it will be termed, henceforth, the PC CARP procedure, or simply PC CARP.

Consider first a simple case where clusters or primary sampling units are taken either with replacement from a finite population or without replacement from a very large finite population. Assume that for the j-th cluster, j = 1, 2, ..., n, a subsample of m_j secondary units is taken either with replacement or with probabilities proportional to some weights. Let y_{k,j}, k = 1, 2, ..., m_j be a (d+1) dimensional vector representing a multinomial variable. This vector consists entirely of zeros except for position i which will contain a one if the observation falls in the i-th cell of the multinomial. Let y_{k,j} represent the vector y_{k,j} without the last category. Associated with y_{k,j} there is a k-dimensional observed vector of explanatory variables x_{k,j}.

For each j = 1, 2, ..., n, and each k = 1, 2, ..., m_j, assume

\[ E(y_{k,j}) = \mu_{k,j} \]

\[ = g(\theta_0, x_{k,j}), \quad (4.6.1) \]
where \( g(z^0, x_j) \) is the generalized logistic function defined in (4.1.6). No assumption on the form of Var(\( y_{x\mid j} \)) is made.

The PC CARP procedure uses algorithm (4.1.12)-(4.1.14) with

\[
\hat{\beta}_n(\hat{y}) = \left( \sum_{j=1}^{n} \sum_{\ell=1}^{m_j} w_{\ell,j} \Delta(\hat{\pi}_{\ell,j}) \right)^{-1} \left( \sum_{j=1}^{n} \sum_{\ell=1}^{m_j} w_{\ell,j}(y_{\ell,j} - \hat{\pi}_{\ell,j}) x_{\ell,j} \right)
\]

where \( w_{\ell,j} \) represents the weight associated with \( y_{\ell,j} \) to estimate \( g^0 \). Let \( \hat{\beta}_{\text{PSEUDO}} \) be the pseudo maximum likelihood estimator resulting from using the modified Gauss-Newton procedure (4.1.12)-(4.1.14). Then, by Theorem (4.2.1)

\[
\sqrt{n}(\hat{\beta}_{\text{PSEUDO}} - g^0) \xrightarrow{d} N_{|G|}(0, A)
\]

where

\[
A = \lim n[H_n(g^0)]^{-1} G_n [H_n(g^0)]^{-1}
\]

almost surely.

\[
H_n(g) = \sum_{j=1}^{n} \sum_{\ell=1}^{m_j} w_{\ell,j} \Delta(g(z^0, x_j)) x_{\ell,j} x_{\ell,j}^t
\]

and

\[
G_n = \sum_{j=1}^{n} \sum_{\ell=1}^{m_j} w_{\ell,j}^2 \text{Var}(y_{\ell,j}).
\]

Observe that a consistent estimator of \( H_n(g^0) \) is \( H_n(\hat{\beta}_{\text{PSEUDO}}) \) and a distribution free estimator of \( G_n \) is

\[
\hat{G}_n = \sum_{j=1}^{n} (d_j - \bar{d})(d_j - \bar{d})^t
\]

(4.6.3)
where

\[ d_j = \sum_{k=1}^{m_j} w_{kj}(y_{kj} - \pi_{kj}) \mathbf{x}_j, \]

and \( \hat{d} = n^{-1} \sum_{j=1}^{n} d_j. \) If within each cluster \( j, \) the \( y_{kj} \)'s are independent and identically distributed according to a multinomial random vector with parameters \((\pi_{kj}, 1)\), then it can be easily shown that expectation of \( \hat{\sigma}_n \) is precisely \( \bar{H}_n(\theta^0). \) In practice the \( \pi_{kj} \)'s in (4.6.3) are replaced with \( \hat{\pi}_{kj} = g(\hat{\theta}_{PSEUDO}, x_{kj}) \) and a small correction is done to the whole covariance matrix. The estimator is

\[ \hat{\Gamma}_n = [(n^* - k)^{-1}(n^* - 1)] \sum_{j=1}^{n} (\hat{d}_j - \hat{d})(\hat{d}_j - \hat{d})', \quad (4.6.4) \]

where

\[ \hat{d}_j = \sum_{k=1}^{m_j} w_{kj}(y_{kj} - \hat{\pi}_{kj}) \mathbf{x}_k, \]

\( \hat{d} = n^{-1} \sum_{j=1}^{n} \hat{d}_j, \) and \( n^* = \sum_{j=1}^{n} m_j. \) Therefore, a consistent estimator of the asymptotic covariance matrix of \( \hat{\theta}_{PSEUDO} \) under the complex sampling design is

\[ \tilde{\Gamma}_n = [H_n(\hat{\theta}_{PSEUDO})]^{-1} \Gamma_n [H_n(\hat{\theta}_{PSEUDO})]^{-1} \quad (4.6.5) \]

which can be used to test any hypothesis of the form \( H_0: C \hat{\theta}^0 = \hat{\xi}. \)

Under the null hypothesis, by Moore (1977)
(4.6.7)

\((C_{\text{PSEUDO}}^{\hat{C}} - \delta^*)'[C_n \tilde{A}_n C'](C_{\text{PSEUDO}}^{\hat{C}} - \delta^*)\) converges in law to a chi-square distribution with \(v = \text{rank } C \tilde{A}_n C'\) degrees of freedom. By analogy to the Hotelling \(T^2\) statistic, it is natural to adjust the test by multiplying (4.6.7) by the ratio

\[\frac{n - v}{v(n - 1)}\]  

(4.6.8)

to obtain an approximate \(F\) statistic with \(v\) and \(n - v\) degrees of freedom. Instead of making this correction in each \(F\), an adjustment that will always produce a positive definite matrix can be made in the covariance matrix. The adjustment is

(1) if \(n > 3dk - 2\)

\[\hat{A}_n = \tilde{A}_n + (n - dk)^{-1}(dk - 1)\gamma^*[H_n(C_{\text{PSEUDO}}^{\hat{C}})]^{-1}\]  

(4.6.9)

(2) if \(n < 3dk - 2\)

\[\hat{A}_n = \tilde{A}_n + 0.5\gamma^*[H_n(C_{\text{PSEUDO}}^{\hat{C}})]^{-1}\]  

(4.6.10)

where \(\gamma^* = \max(1, \text{tr}([H_n(C_{\text{PSEUDO}}^{\hat{C}})]^{-1}\tilde{C}_n)/dk)\). The upper bound of 0.5 for the correction in (4.6.10) is arbitrary. An approximate \(F\) test with \(v\) and \(n - v\) degrees of freedom is obtained by substituting \(\hat{A}_n\) for \(\tilde{A}_n\) in (4.6.7) and dividing the resulting quadratic form by \(v\).
In practice, PC CARP assumes that \( n \) is large so the approximate
degrees of freedom are \( v \) and infinity.

An estimator of \( \phi^0 \) of Model II in Section 4.2 can be obtained as

\[
\hat{\phi} = (dk)^{-1} \sum_{i=1}^{dk} \hat{a}(i,i)/\hat{h}(i,i)
\]

(4.6.11)

where \( \hat{a}(i,i) \) and \( \hat{h}(i,i) \) represent the \((i,i)\)-th elements of \( \hat{A}_n \) and
\( [H_n(\hat{B}_{PSEUDO})]^{-1} \), respectively. The estimator \( \hat{\phi} \) can be viewed as an
estimator of the average design effect.

A generalization of PC CARP procedure to more complicated sampling
designs can be done as follows. Suppose that the whole population has
been divided into \( s = 1, 2, \ldots, S \) strata. Let \( y_{ljs} \) denote the \( l \)-th
multinomial vector, \( l = 1, 2, \ldots, m_{js} \) within the \( j \)-th primary, \( j = 1, 2, \ldots, n_s \), in the \( s \)-th stratum, \( s = 1, 2, \ldots, S \). Assume that

\[
E(y_{ljs}) = \pi_{ljs}
\]

\[
= g(\theta^0, x_{ljs}).
\]

(4.6.12)

Then, the algorithm (4.1.13)-(4.1.15) is computed using

\[
\hat{\sigma}_n(\hat{g}) = [\sum_{s=1}^{S} \sum_{j=1}^{n_s} \sum_{l=1}^{m_{js}} w_{ljs} \Delta(\hat{\pi}_{ljs}) \times x_{ljs} x'_{ljs}]^{-1}
\]

\[
\times [\sum_{s=1}^{S} \sum_{j=1}^{n_s} \sum_{l=1}^{m_{js}} w_{ljs}(y_{ljs} - \hat{\pi}_{ljs}) \times x_{ljs}]
\]

(4.6.13)
where the weights are

\[ w_{js} = (f_s^{-1}) (F_s^{-1}) \]

\( f_s \) is the primary sampling unit rate within the stratum and \( F_s \) is the subsampling rate for elements within primary sampling units. The arguments used before also apply here. The covariance matrix \( \hat{A}_n \) is computed following the rules given in (4.6.9)-(4.6.10) with

\[
H_n^{(\hat{\sigma}_N \text{ PSEUDO})} = \sum_{s=1}^{S} \sum_{j=1}^{n_j} \sum_{l=1}^{m_{sj}} w_{js} \Delta(\hat{\sigma}_{js} \Delta \hat{\sigma}_{js}) = x_{js} x'_{js} \tag{4.6.14}
\]

and

\[
\hat{G}_n = (n^*-k)^{-1}(n^*-1) \sum_{s=1}^{S} (n_s - 1)^{-1} n_s (1 - f_s) \sum_{j=1}^{n_s} (\hat{d}_{js} - \bar{d}) (\hat{d}_{js} - \bar{d}), \tag{4.6.15}
\]

where

\[
\hat{d}_{js} = \sum_{l=1}^{m_{js}} w_{ljs} (y_{ljs} - \hat{\mu}_{ljs}) = x_{ljs},
\]

\[
\hat{d} = n^{-1} \sum_{j=1}^{n_s} \hat{d}_{js},
\]

and

\[
n^* = \sum_{s=1}^{S} \sum_{j=1}^{n_j} m_{sj}.
\]
4.7. An Example

In this section an example is presented to illustrate some of the models described before. A simple random sample of size $n = 2,500$ households was taken from another sample of $n' = 58,000$ households. Each household represents a cluster and households are assumed to be independent in the sample of $n = 2,500$. The households were classified into two types with respect to their location. There are $n_1 = 1,631$ urban households and $n_2 = 869$ rural households in the sample. Information on the weekly household expenditures in certain categories was collected for each household. In addition to this, information on some covariates was also collected. Initially, separate logit models are fit to the rural and urban households. A description of the variables used for each case is provided in Table 4.1. In defining the generalized logistic function (4.1.6), the category of weekly expenditure on nonfoods will be treated as the last category. For each of the first four categories there are six beta coefficients to be estimated so, the total number of parameters to be estimated is 24.

The models examined here are extra-variation generalizations of models proposed by Theil (1969). Let $\pi_i$ represent the proportion of income spent on the $i$-th commodity. Theil's model describes each $\pi_i$ in terms of covariates $x$ (income, family size, etc.) and $y$ (prices of commodities) as
Table 4.1. Description of the variables for each household

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>Weekly proportion spent on rice</td>
</tr>
<tr>
<td>$P_2$</td>
<td>Weekly proportion spent on fruits and vegetables</td>
</tr>
<tr>
<td>$P_3$</td>
<td>Weekly proportion spent on meats, dairy and fish</td>
</tr>
<tr>
<td>$P_4$</td>
<td>Weekly proportion spent on miscellaneous foods</td>
</tr>
<tr>
<td>$P_5$</td>
<td>Weekly proportion spent on nonfoods</td>
</tr>
<tr>
<td>$x_1$</td>
<td>One if household is urban, zero otherwise</td>
</tr>
<tr>
<td>$x_2$</td>
<td>Log of total household income if household is urban, zero otherwise</td>
</tr>
<tr>
<td>$x_3$</td>
<td>Log of people in the household if household is urban, zero otherwise</td>
</tr>
<tr>
<td>$x_4$</td>
<td>One if household is rural, zero otherwise</td>
</tr>
<tr>
<td>$x_5$</td>
<td>Log of total household income if household is rural, zero otherwise</td>
</tr>
<tr>
<td>$x_6$</td>
<td>Log of people in the household if household is rural, zero otherwise</td>
</tr>
</tbody>
</table>

\[
\ln\left(\frac{\pi_i}{\pi_{d+1}}\right) = \beta_0 + \mathbf{x}' \beta_i + y' \chi, \quad i = 1, 2, \ldots, d \quad (4.7.1)
\]

which is equivalent to

\[
\pi_i = [1 + \sum_{\ell=1}^{d} \exp(\beta_0 + \mathbf{x}' \beta_\ell + y' \chi)]^{-1} \exp(\beta_0 + \mathbf{x}' \beta_1 + y' \chi), \ldots, \\
\exp(\beta_0 + \mathbf{x}' \beta_d + y' \chi)] . \quad (4.7.2)
\]
In this particular example \( x = (X_1, X_2, ..., X_6)' \), where \( X_1, X_2, ..., X_6 \) are defined in Table 4.1. There is no information available on \( y \), the prices of the commodities.

The first model considered is the logistic multinomial model described in Section 4.1, which is essentially Theil's model. It assumes that each unit of money is allocated independently to each of the five expenditure categories, which may be quite unrealistic. The beta estimates were computed using the algorithm described in (4.1.12)-(4.1.14). Next, the model described in Section 4.2 was fit. It provides the same estimates of the beta coefficients as those computed under the previous model. Finally, the model described in Section 4.3 was also fit. Two cases were considered: one when \( \beta^0 \) is diagonal and the other one when \( \beta^0 \) is completely unspecified.

Let \( \hat{\beta} \) be the estimated vector of coefficients of the logistic model when any of the four models is fit, \( \hat{\beta} = \text{Vec}(\hat{\beta}_1, ..., \hat{\beta}_4) \), \( \hat{\beta}_i = (\hat{\beta}_{i1}, ..., \hat{\beta}_{i6})' \), \( i = 1, ..., 4 \). Then

\[
\sqrt{n}(\hat{\beta} - \beta^0) \xrightarrow{d} N_{24}(0, \Sigma^{-1})
\]

where \( \beta^0 \) is the true parameter vector of the logistic model. If one wants to test the hypothesis \( H_0: C\beta^0 = 0 \) where \( C \) is an \( r \times 24 \) matrix of rank \( v \), then by Moore (1977)

\[
n[C(\hat{\beta} - \beta^0)'][C\Sigma^{-1}_n C'][C(\hat{\beta} - \beta^0)] \xrightarrow{d} \chi^2(v)
\]
where \( \hat{V}_n^{-1} \) is any consistent estimator of \( V^{-1} \) and \( [C\hat{V}_n^{-1}C']^{-1} \) is any generalized inverse of \( [C\hat{V}_n^{-1}C'] \). Then by proper choice of \( C \), several asymptotic chi-square tests can be computed under the maintained assumption that the model specification is true. Table 4.2 provides a description of some \( C \) matrices needed for testing certain hypotheses of the form \( H_0: C \beta^0 = 0 \). In Table 4.3 the values of each of the four hypotheses described in Table 4.2 are shown for the four models being fit. Observe that the asymptotic chi-square tests are very large for all the hypotheses being tested under the logistic multinomial model.

The assumption that in each household each dollar is allocated independently into one of the five expenditure categories produces a very small the standard error of the beta estimates and, consequently, very large chi-square tests. The other four models are extra-variability models.

Table 4.2. Selection of \( C \) matrix for testing the hypothesis

\( H_0: C \beta^0 = 0 \)

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>( C ) matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>The coefficients for &quot;Intercept, Log Income and Log of Number of People&quot; are the same for urban and rural households.</td>
<td>( C_1 = \frac{1}{4} \left(I_3, -I_3\right) )</td>
</tr>
<tr>
<td>The coefficients for &quot;Intercept&quot; are the same for urban and rural households.</td>
<td>( C_2 = \frac{1}{4} \left(1, 0, 0, -1, 0, 0\right) )</td>
</tr>
<tr>
<td>The coefficients for &quot;Log Income&quot; are the same for urban and rural households.</td>
<td>( C_3 = \frac{1}{4} \left(0, 1, 0, 0, -1, 0\right) )</td>
</tr>
<tr>
<td>The coefficients for &quot;Log Number of People&quot; are the same for urban and rural households.</td>
<td>( C_4 = \frac{1}{4} \left(0, 0, 1, 0, 0, -1\right) )</td>
</tr>
</tbody>
</table>
Table 4.3. Values of the chi-square tests for certain hypotheses of the form \( H_0: Cg^0 = 0 \) under alternative models

<table>
<thead>
<tr>
<th>Hypothesis ( H_0: Cg^0 = 0 )</th>
<th>Model</th>
<th>( C = C_1 )</th>
<th>( C = C_2 )</th>
<th>( C = C_3 )</th>
<th>( C = C_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Multinomial</td>
<td>135020**</td>
<td>11715**</td>
<td>2118**</td>
<td>744**</td>
<td></td>
</tr>
<tr>
<td>Scaled Multinomial Variance</td>
<td>286**</td>
<td>24.79**</td>
<td>4.48*</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>Generalized Multinomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance with ( \xi^0 ) diagonal</td>
<td>307**</td>
<td>27.07**</td>
<td>4.99*</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>Generalized Multinomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance with ( \xi^0 ) unspecified</td>
<td>240**</td>
<td>20.72**</td>
<td>3.81</td>
<td>1.39</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at 0.05.
**Significant at 0.01.

The standard errors computed under the extra variability models are much bigger than those obtained under the logistic multinomial model. This is reflected in the fact that the chi-square statistics are between 429 and 565 times smaller than those computed under the logistic multinomial model. Inferences made under the extra variability models, are in some cases completely different from the conclusions drawn under the logistic multinomial model. For instance, the null hypothesis \( H_0: C_4g^0 = 0 \) is not rejected at 5% in the extra variability models but is rejected under the the multinomial model. In the case of \( H_0: C_3g^0 = 0 \) the chi-square test is barely significant at 5% for the second and third models and not significant for the last model. Some conclusions can be derived from Table 4.3. It seems to be reasonable to reject the hypotheses.
H₀: C_iβ^0 = \mathbf{0} \text{ for } i=1, 2, 3. This suggests that separate models must be fit to the urban and rural households.

Now, consider only the sample of \( n_1 = 1,631 \) urban household and let \( \mathbf{x} = (x_1, x_2, x_3)' \). For this particular case \( \hat{\beta}_i \) is a column vector of dimension 12 such that \( \hat{\beta}_i = \text{vec}(\beta_{i1}, \ldots, \beta_{i4}) \), \( \hat{\beta}_{i1} = (\hat{\beta}_{i1}, \hat{\beta}_{i2}, \hat{\beta}_{i3})' \), \( i=1, \ldots, 4 \). The four models described previously were fit to the urban household data. The results of the logistic multinomial model are reported in Table 4.4. Both the beta estimates and standard errors were multiplied by ten.

<table>
<thead>
<tr>
<th></th>
<th>Rice</th>
<th>Vegetables and fruits</th>
<th>Meats, dairy, and fish</th>
<th>Miscellaneous foods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.370</td>
<td>0.318</td>
<td>-5.432</td>
<td>9.426</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.049)</td>
<td>(0.042)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Log of Income</td>
<td>-9.011</td>
<td>-5.166</td>
<td>-0.709</td>
<td>-7.123</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Log No. of People</td>
<td>7.925</td>
<td>2.200</td>
<td>0.144</td>
<td>4.001</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.022)</td>
<td>(0.019)</td>
<td>(0.018)</td>
</tr>
</tbody>
</table>

For the second model, the estimator (4.2.8) of the extra parameter \( \phi^0 \) in the scaled multinomial variance model is \( \hat{\phi} = 509.60 \). The beta estimates are exactly the same as those computed under the logistic multinomial model, but the standard errors are 22.6 times greater than
those computed under the multinomial model. The results are shown in Table 4.5.

Table 4.5. Parameter estimates and standard errors under the scaled multinomial variance model

<table>
<thead>
<tr>
<th></th>
<th>Rice</th>
<th>Vegetables and fruits</th>
<th>Meats, dairy, and fish</th>
<th>Miscellaneous foods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.370</td>
<td>0.318</td>
<td>-5.432</td>
<td>9.426</td>
</tr>
<tr>
<td></td>
<td>(0.808)</td>
<td>(1.098)</td>
<td>(0.946)</td>
<td>(0.889)</td>
</tr>
<tr>
<td>Log of Income</td>
<td>-9.011</td>
<td>-5.166</td>
<td>-0.709</td>
<td>-7.123</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.403)</td>
<td>(0.335)</td>
<td>(0.328)</td>
</tr>
<tr>
<td>Log No. of People</td>
<td>7.925</td>
<td>2.200</td>
<td>0.144</td>
<td>4.001</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.500)</td>
<td>(0.424)</td>
<td>(0.407)</td>
</tr>
</tbody>
</table>

Finally, the last two models described in Section 4.3 are fit: the first one when $\phi^0$ is diagonal and the other one when $\phi^0$ is completely unspecified. The estimates of the extra parameters are:

$$\hat{\phi} = \text{Diag}(595.77, 338.01, 633.26, 471.39)$$  \hspace{1cm} (4.7.3)

when $\phi^0$ is assumed diagonal and,

$$\hat{\Sigma} = \begin{pmatrix} 595.77 & \text{symmetric} \\ 49.960 & 338.01 \\ 33.483 & 80.211 & 633.26 \\ -70.456 & 116.59 & 72.272 & 471.39 \end{pmatrix}$$  \hspace{1cm} (4.7.4)
when $\psi^0$ is completely unspecified. The estimates of the beta coefficients and standard errors are presented in Tables 4.6 and 4.7. As before, the original results were multiplied by ten.

Table 4.6. Parameter estimates and standard errors under the generalized multinomial variance model with $\psi^0$ diagonal

<table>
<thead>
<tr>
<th></th>
<th>Rice</th>
<th>Vegetables and fruits</th>
<th>Meats, dairy, and fish</th>
<th>Miscellaneous foods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.560</td>
<td>0.188</td>
<td>-5.150</td>
<td>9.372</td>
</tr>
<tr>
<td></td>
<td>(0.863)</td>
<td>(0.909)</td>
<td>(1.047)</td>
<td>(0.861)</td>
</tr>
<tr>
<td>Log of Income</td>
<td>-9.106</td>
<td>-5.065</td>
<td>-0.852</td>
<td>-7.081</td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
<td>(0.333)</td>
<td>(0.371)</td>
<td>(0.317)</td>
</tr>
<tr>
<td>Log No. of People</td>
<td>7.978</td>
<td>2.085</td>
<td>0.229</td>
<td>3.954</td>
</tr>
<tr>
<td></td>
<td>(0.402)</td>
<td>(0.413)</td>
<td>(0.469)</td>
<td>(0.393)</td>
</tr>
</tbody>
</table>

Table 4.7. Parameter estimates and standard errors under the generalized multinomial variance model with $\psi^0$ completely unspecified

<table>
<thead>
<tr>
<th></th>
<th>Rice</th>
<th>Vegetables and fruits</th>
<th>Meats, dairy, and fish</th>
<th>Miscellaneous foods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.482</td>
<td>0.357</td>
<td>-5.226</td>
<td>9.453</td>
</tr>
<tr>
<td></td>
<td>(0.879)</td>
<td>(0.983)</td>
<td>(1.092)</td>
<td>(0.895)</td>
</tr>
<tr>
<td>Log of Income</td>
<td>-9.083</td>
<td>-5.157</td>
<td>-0.829</td>
<td>-7.129</td>
</tr>
<tr>
<td></td>
<td>(0.323)</td>
<td>(0.359)</td>
<td>(0.386)</td>
<td>(0.329)</td>
</tr>
<tr>
<td>Log No. of People</td>
<td>7.985</td>
<td>2.148</td>
<td>0.235</td>
<td>3.989</td>
</tr>
<tr>
<td></td>
<td>(0.408)</td>
<td>(0.446)</td>
<td>(0.489)</td>
<td>(0.408)</td>
</tr>
</tbody>
</table>
The beta estimates are very similar to those obtained under the logistic multinomial model, except for the estimate of the intercept for the category "Vegetables and Fruits" for the model with $\delta^0$ diagonal. The standard errors in these last two models are roughly 20 times bigger than those under the logistic multinomial model. The estimates of the standard errors are very similar for the last three models.
5. A MONTE CARLO STUDY

In this chapter a Monte Carlo study is conducted to examine the properties of chi-square tests involving model parameters. Some of the estimation procedures outlined in Chapter IV are applied to data generated under two different sampling schemes. Properties of estimators for the extra variation parameter and of other parameter estimators for logistic models are further investigated.

5.1. Data Generation Under Two Sampling Schemes

The types of data generated in this Monte Carlo study correspond to two different sampling schemes: logistic multinomial data satisfying Model I in Section 4.1 and, data with extra multinomial variation in logistic models satisfying Model II in Section 4.2. The latter type of data are obtained by generating multinomial random vectors within each cluster with a specific intra-class correlation.

Let \( x_1, x_2, \ldots, x_n \) be a sample of size \( n \) of \( k \)-dimensional random vectors from a population normally distributed with mean vector \( \mu \) and covariance matrix \( \Sigma \). Let \( \beta_0^1, \beta_0^2, \ldots, \beta_0^d \) be a set of known \( k \)-dimensional vectors. For each \( j=1, 2, \ldots, n \), define \( y_j \) to be a \( d \)-dimensional random vector such that

\[
y_j = (x_{j1} \beta_1^0, x_{j2} \beta_2^0, \ldots, x_{jd} \beta_d^0)'.
\]

Then \( y_j \) is distributed as a \( d \)-dimensional normal random variable with
mean vector

\[ \mu^* = (\mu_1^0, \mu_2^0, \ldots, \mu_d^0)' \]  
\hspace{0.5cm} (5.1.2)

and covariance matrix

\[ \Sigma^* = [\Sigma_1^0, \Sigma_2^0, \ldots, \Sigma_d^0]' [\Sigma_1^0, \Sigma_2^0, \ldots, \Sigma_d^0] . \]  
\hspace{0.5cm} (5.1.3)

Therefore,

\[ \pi_j = \left[ 1 + \sum_{i=1}^{d} \exp(y_{ij}) \right]^{-1} \left[ \exp(y_{1j}), \ldots, \exp(y_{dj}) \right]' , \]  
\hspace{0.5cm} (5.1.4)

where \( y_{ij} = \mathbf{x}_j^0 \mathbf{y}_i^0 \), \( i=1, 2, \ldots, d \), has a logistic-normal distribution with parameters \( (\mu^*, \Sigma^*) \). Properties and applications of the logistic-normal distribution can be found in Aitchinson and Shen (1980). The parameters \( (\mu^*, \Sigma^*) \) do not represent the mean vector and covariance matrix of the logistic-normal distribution. The mean vector and covariance matrix of the logistic normal do not have closed form representation, unlike the Dirichlet-multinomial distribution [see for example Mosimann (1962)]. The Dirichlet-multinomial has been used to generate vectors of probabilities, but the logistic-normal distribution seems to be more realistic than the Dirichlet-multinomial because the logistic-normal generates probability vectors with more extreme values and because the logistic-normal allows the expectation of the observed vector of probabilities \( \pi_j \), to be linked to a reduced set of
covariates. Aitchinson and Begg (1976) indicated that any Dirichlet distribution can be closely approximated by a suitable logistic-normal distribution.

In this study, the \( \pi_j \)'s are generated according to a logistic-normal distribution with parameters \((\mu^*, \xi^*)\) in (5.1.2) and (5.1.3), where

\[
\mu = (1, -2, 0, 5),
\]

\[
\xi = \begin{pmatrix}
0 & 0 \\
0 & 25I_3
\end{pmatrix},
\]

\[
\xi_0 = (-0.3, -0.1, 0.1, 0.3)',
\]

\[
\xi_1 = (0.2, -0.2, -0.2, 0.2)',
\]

and

\[
\xi_3 = (-0.1, 0.3, -0.3, 0.1).
\]

The number of categories is \( d+1 = 4 \), and the number of covariates or explanatory variables is \( k = 4 \) with the first element of each \( x_j \) being always one, i.e., an intercept has been included. For generating the \( x_j \)'s the method of Box and Muller (1958) was implemented. The uniform random digits were provided by the algorithm for generating pseudo-random numbers proposed by Wichmann and Hill (1982). Their
algorithm uses three simple multiplicative congruential generators and it has a cycle length exceeding \( 2.70 \times 10^{13} \).

Some Monte Carlo properties of the true probability vectors \( \pi_j \)'s are presented in Appendix A. The properties are based on a sample of 10,000 logistic-normal random vectors (5.1.4) with parameters (5.1.2), (5.1.3), (5.1.5)-(5.1.9).

Once \( \pi_j \) is generated, it is fixed, and a set \( y_{1,j}, y_{2,j}, \ldots, y_{m_j,j} \) of \( d \)-dimensional random vectors is generated using two different sampling schemes. In the first scheme the \( y_{1,j}, y_{2,j}, \ldots, y_{m_j,j} \) are independent and identically distributed according to a multinomial distribution with parameters \((1, \pi_j)\).

Then, given \( x_j \),

\[
m_j \mathbf{P}_j = \sum_{k=1}^{m_j} y_{k,j} \sim \text{Multinomial} (m_j, \pi_j), \quad (5.1.10)
\]

therefore, any sequence of data \( \{(\mathbf{P}_j, x_j)_{j=1}^n\} \) satisfies Model I of Section 4.1.

If for a given \( \pi_j \), the \( y_{k,j}, k=1, 2, \ldots, m_j \) are generated such that

\[
y_{k,j} = \begin{cases} 
y_{0,j} & \text{with probability } \zeta \\
y_{k,j}^* & \text{with probability } 1 - \zeta 
\end{cases}
\]
where \( y_{0j}, y_{1j}, \ldots, y_{mj}, \ldots \) are independent and identically distributed Multinomial \((1, \pi_j)\) random variable, then it can be shown that within the \( j \)-th cluster

\[
E(y^j_{k_j}) = \pi_j
\]

(5.1.11)

and

\[
\text{Cov}(y^j_{k_j}, y^j_{s_j}) = \zeta^2 A(\pi_j), \quad j \neq s.
\]

(5.1.12)

Therefore, given \( \pi_j \), \( m_j \pi_j = \sum_{j=1}^{m_j} y^j_{k_j} \) does not have a multinomial distribution. Instead

\[
E(P_j) = \pi_j
\]

(5.1.13)

and

\[
\text{Var}(P_j) = [1 + (m_j - 1)\zeta^2]m_j^{-1}A(\pi_j),
\]

(5.1.14)

where \( \zeta \) is an extra parameter that represents the intra-class correlation in cluster sampling. Data \( \{P_j, x_j\}_{j=1}^{n} \) generated under this mechanism represent the multinomial generalization of binomial data with extra binomial variation analyzed by Williams (1982). Furthermore, if the \( m_j \)'s are taken to be constant, i.e., \( m_j = m \), then the factor \( [1 + (m-1)\zeta^2] \) can be viewed as the extra parameter \( \phi^0 \) of Section 4.2. In such a case
\[ \text{Var}(P_j) = \phi^0 m^{-1} \Delta(x_j), \]  
\[ (5.1.15) \]

where \( \phi^0 = [1 + (m-1)s^2] \).

A congruential generator of period \( 2^{31} - 2 \), belonging to the family of congruential generators discussed by Kennedy and Gentle (1980) was used for generating the \( y_{k_j} \)'s. Different combinations of \( n \), \( m \) and \( \phi^0 \) were set in advance. For each combination, 1000 sets of samples \( \{(P_j, x_j)_{j=1}^n\} \) were generated according to the logistic multinomial model if \( \phi^0 = 1 (\zeta = 0) \), or according to the scaled multinomial variance model if \( \phi^0 > 1 (\zeta > 1) \). The seeds for the two uniform random generators were saved in a file. Then, at the end of each run of 1000 sets of samples \( \{(P_j, x_j)_{j=1}^n\} \), the final seeds were saved in the same file and used as initial seeds for the next run. In this way, the pseudo-random sequences used in different runs were never duplicated.

5.2. Computed Statistics

For each set of 1000 samples \( \{(P_j, x_j)_{j=1}^n\} \) generated at each of the different combinations of \( n \), \( m \) and \( \phi^0 \) used in this Monte Carlo study, the following statistics were computed:

1) \( \hat{\theta} = \text{vec}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) \), an estimator of the parameters (5.1.7)-(5.1.9) of the logistic model. This estimator was computed using the modified Gauss-Newton algorithm (4.1.12)-(4.1.14). In order to obtain a more accurate solution of the maximum likelihood equations, the stopping rule (4.1.14) was changed to

\[ \]
\[
\frac{x(\hat{g}^{(i*)}) - x(\hat{g})}{|x(\hat{g}^{(i*)})| + 10^{-6}} < 10^{-8} . \tag{5.2.1}
\]

2) The estimated covariance matrix of \( \hat{g} \) obtained from the three procedures: logistic multinomial of Section 4.1, scaled multinomial variance of Section 4.2 and PC CARP of Section 4.6. Then, chi-square tests of \( H_0: \hat{g} = \hat{g}^0 \) were computing using the quadratic form

\[
(\hat{g} - \hat{g}^0)'[\text{Cov}(\hat{g})]^{-1}(\hat{g} - \hat{g}^0) \tag{5.2.2}
\]

for each of these estimation procedures. The Type I error with nominal 0.05 level was estimated along with the percentiles of the empirical distributions of the statistic (5.2.2). In addition, "t" statistics for the individual coefficient estimates were also computed. These individual test statistics are defined as

\[
\text{"t"}_{\hat{g}} = [\text{Var}(\hat{g}_{iZ})]^{-1/2}(\hat{g}_{iZ} - \hat{g}_{iZ}^0) , \tag{5.2.3}
\]

where \( i=1, 2, 3 \) and \( Z=1, 2, 3, 4 \). The twelve "t" statistics provided by the estimated parameter for each of the three estimation procedures were grouped together and the percentiles computed. Consequently, for each run the percentiles are based on 12,000 "t" values.

3) The vector of biases \( \hat{g} - \hat{g}^0 \) and Monte-Carlo "t" statistics for testing if the individual biases are significantly different from zero. The nominal level was 0.05. Observe that the biases under the
three estimation procedures used here are the same because the estimators are the same. Only the variance estimation methods differ.

4) The estimator \( \hat{\phi} \) of equation (4.2.8) and \( \tilde{\phi} \) of equation (4.6.10). The Monte Carlo mean and standard deviations are reported.

5) The one-step Gauss-Newton estimator \( \hat{\beta}^{(1)} \) for the generalized multinomial variance procedure of Section 4.3 with \( \text{Diag} (\theta^0) \). The quadratic form (5.2.2) was also computed using the proper estimated covariance matrix of \( \hat{\beta}^{(1)} \) under this procedure. As before, the Type I error with nominal 0.05 level and the percentiles of the asymptotically chi-square statistic were estimated. The "t" values defined in (5.2.3) were grouped together in the estimation of the percentiles for the "t" statistics.

6) The estimators for the extra parameter \( \text{Diag} \phi^0 \) described in (4.3.3)-(4.3.4). Monte Carlo means and standard deviations are reported.

Note that the adjustment in (4.6.8)-(4.6.10) has been done under the assumption for each \( j \), \( j=1, 2, \ldots, n \), the variables \( x_{ij}, \ldots, x_{m,j} \) are not the same. If within each cluster the x's are the same, an adjustment like the one in (4.6.8)-(4.6.10) should be done replacing (4.6.8) with

\[
\frac{n - k - v - 1}{v(n - k)}.
\]

(5.2.4)

In this case, the second term in the right hand side of (4.6.9) is
\[
\frac{dk + 1}{n - k - dk + 1} \nu \left[ \mathbb{H} \left( \hat{\theta}_{\text{PSEUDO}} \right) \right]^{-1} \tag{5.2.5}
\]

and the upper bound of 0.5 for the correction in (4.6.10) is done if

\[ n < 3dk + 3 + k . \tag{5.2.6} \]

In this simulation the \( \mathbf{x}'s \) within each cluster are the same, so correction (5.2.4)-(5.2.6) should be implemented. However, since PC CARP has been designed to operate under the assumption of different \( \mathbf{x}'s \) within each cluster, the covariance matrix of \( \hat{\theta} \) will be computed using (4.6.8)-(4.6.10).

5.3. Results

The different combinations of \( n, m, \phi^0 \) and \( \zeta \) used in the Monte Carlo study are shown in Table 5.1 along with the average numbers of steps needed for the convergence criterion to be met. As stated in Section 4.1 a modified Gauss-Newton procedure should always converge to the maximum likelihood estimator of \( \hat{\theta}^0 \) when \( \{(\mathbf{P}_j, \mathbf{x}_j)^n_{j=1}\} \) satisfies the logistic multinomial model of Section 4.1. Haberman (1974) proved that this convergence takes place independently of the initial solution. In this simulation, two different initial solutions were used: the weighted least squared estimator (4.1.9)-(4.1.11), denoted by WLS, and the starting value where all the parameters are set to be zero, which has been denoted by \( \theta = 0 \). The convergence, as one can see in Table 5.1 was met in approximately four steps when the WLS initial
Table 5.1. Average number of steps until convergence was met for different values of \( n \), \( m \), \( \phi^0 \), \( \zeta \) and different initial solutions

| \( n \) | \( m \) | \( \phi^0 \) | \( \zeta \) | Initial solution | Average number of steps until convergence was met |
|--------|--------|*********|***********|-----------------|-----------------------------------------------|
| 30     | 28     | 1        | 0         | \( \hat{\beta} = 0 \) | 5.89                                          |
| 100    | 1      | 1        | 0         | \( \hat{\beta} = 0 \) | 6.23                                          |
| 100    | 28     | 1        | 0         | WLS             | 4.00                                          |
| 300    | 28     | 1        | 0         | WLS             | 4.00                                          |
| 30     | 26     | 2        | 1/5       | WLS             | 4.15                                          |
| 100    | 5      | 2        | 1/2       | \( \hat{\beta} = 0 \) | 5.91                                          |
| 100    | 26     | 2        | 1/5       | WLS             | 4.01                                          |
| 300    | 26     | 2        | 1/5       | WLS             | 4.00                                          |
| 30     | 28     | 4        | 1/3       | WLS             | 4.25                                          |
| 100    | 28     | 4        | 1/3       | WLS             | 4.02                                          |
| 300    | 28     | 4        | 1/3       | WLS             | 4.00                                          |

solution is used. A couple of more steps were necessary when the initial parameter vector was \( \hat{\beta} = 0 \). The initial solution WLS defined in (4.1.9)-(4.1.11) and one step of the Gauss-Newton procedure require essentially the same amount of computation. However, it seems the iterative procedure converges faster, in cases where WLS estimators are feasible. This can be inferred from the cases with \( n = 30 \).

The estimated Type I errors obtained from comparing the chi-square tests of \( H_0: \hat{\beta} = \beta_0 \) against \( \chi^2(12, 0.05) = 21.03 \) are presented in Table 5.2 under four different estimation procedures: logistic multinomial variance, scaled multinomial variance, generalized
Table 5.2. Estimated Type I error for the chi-square test of $H_0: \beta = \beta^0$ when $\phi^0 = \phi^*$ with nominal 0.05 level

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>$\phi^*$</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>0.05</td>
<td>0.08</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>0.03</td>
<td>---</td>
<td>---</td>
<td>0.06</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>0.05</td>
<td>0.07</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>0.53</td>
<td>0.15</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>0.54</td>
<td>0.12</td>
<td>0.15</td>
<td>0.12</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>0.57</td>
<td>0.09</td>
<td>0.10</td>
<td>0.12</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>0.56</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>0.94</td>
<td>0.21</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>0.95</td>
<td>0.11</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>0.95</td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
</tr>
</tbody>
</table>

multinomial variance (with $\text{Diag}\ \phi^0$) and single stage cluster analysis in PC CARP.

For data generated under the logistic multinomial model ($\phi^0 = 1$) the nominal 5% level is quite accurately achieved for the logistic multinomial variance procedure except for the case when $n = 100$, $m = 1$, where the nominal level is underestimated by 2%. Observe that for the case $n = 30$, $m = 28$, the number of independent multinomial realizations is 8.4 times greater than when $n = 100$, $m = 1$, so, it seems that the 2% underestimation is related to the sample size. The estimated size of the scaled multinomial variance procedure as well as
of the generalized multinomial one, exceed the nominal 5% level, the estimated level approaches the effective nominal level as the sample size increases. Note that for the case $n = 300$, $m = 28$ the nominal level is only slightly smaller than the observed level. PC CARP procedure produces an estimated Type I error that differs from the nominal one by only 1% when $n = 30$, $m = 28$, and $n = 100$, $n = 1$. Because of the standard covariance matrix (4.6.9)-(4.6.10), the estimated nominal level goes up for $n = 100$, $m = 28$ and then goes down to 7% when $n = 300$, $m = 28$. It seems that PC CARP works fairly well for small samples, then Type I error levels become more inflated when the sample size increases, but eventually gets closer to the nominal level as the sample size keeps increasing.

There is a strong distortion of the estimated Type I error for the logistic multinomial variance procedure when extra variability (or intra-class correlation) is present. The estimated nominal 0.05 level varies from 0.53 to 0.57 when $\phi^0 = 2$ and, from 0.94 to 0.95 when $\phi^0 = 4$. The other three procedures perform much better. However, for $n = 30$, $m = 28$, $\phi^0 = 4 (\zeta = 1/3)$ the estimated level is still large. The level improves when the sample size increases, as expected. For instance, in the case $n = 300$, $m = 26$, $\phi^0 = 2 (\zeta = 1/5)$ the estimated Type I error varies from 0.06 to 0.08 for the three estimation procedures that recognize extra variation. If the extra variability increases, i.e. $\phi^0 = 4$, larger samples are needed to reduce Type I error levels to the nominal 0.05 level. When $n = 300$, $m = 28$, $\phi^0 = 4$ the estimated level varies from 9% to 10%. 
The percentiles for the chi-square test of \( H_0: \theta = \theta^0 \) in (5.2.2) were calculated for the four estimation procedures and for the combinations of \( n \), \( m \) and \( \phi^0 \) that appear in Table 5.2. Since the 95-th percentile and the Type I error with nominal 0.05 level show the same results, no comments will be made on the 95-th percentiles. The 99-th, 95-th, 90-th, 75-th and 50-th percentiles can be found in Appendix B.

The Monte Carlo means and standard deviations for the estimators \( \hat{\phi} \) (equation 4.2.9) and \( \tilde{\phi} \) (equation 4.6.11) of \( \phi^0 \) are presented in Table 5.3. For \( \phi^0 = 1 \), the estimator \( \hat{\phi} \) has Monte Carlo mean very close to one, while \( \tilde{\phi} \) overestimates \( \phi^0 \), especially for the case

<table>
<thead>
<tr>
<th>Statistic</th>
<th>( \hat{\phi} ) (eqn. 4.2.9)</th>
<th>( \tilde{\phi} ) (eqn. 4.6.11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monte Carlo mean</td>
<td>Monte Carlo s.d.</td>
</tr>
<tr>
<td>( n )</td>
<td>( m )</td>
<td>( \phi^0 )</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
</tr>
</tbody>
</table>
n = 30, m = 28. With no extra variability both estimators have essentially the same Monte Carlo standard deviations.

In cases where $\phi^0 = 2$ or $\phi^0 = 4$, the Monte Carlo means of $\hat{\phi}$ show an underestimation of $\phi^0$. It seems that this negative bias becomes more pronounced as $\phi^0$ increases when $n$ and $m$ are held constant. On the other hand, if $\phi^0$ is held constant the bias decreases as the sample size increases. These relationships hold for $\tilde{\phi}$ as well. It seems that the bias is a function of $n^{-1}$ and possibly of $n^{-1}\phi^0$, so it could be estimated from $n^{-1}$ and $n^{-1}\phi$. In the presence of extra variability, the estimator $\tilde{\phi}$ tends to produce a bias smaller than that incurred by using $\hat{\phi}$.

When $\phi^0 = 4$, the estimator $\hat{\phi}$ seems to have a variability that does not decrease as the sample increases. Its Monte Carlo standard error decreases from 1.18 when $n = 30, \phi^0 = 4$ to 1.12 when $n = 100, \phi^0 = 4$ and then increases to 1.92 when $n = 300, \phi^0 = 4$. It was found that in this last case the $\hat{\phi}$'s ranged from 2.46 to 60.62. This Monte Carlo standard deviation of 1.92 is due to the maximum value of the $\hat{\phi}$'s 60.62. If that value is deleted the standard deviation computed from the remaining 999 estimates is 0.69. An additional simulation for this case provided a Monte Carlo standard deviation of 0.67 and a Monte Carlo mean of 3.81. This shows that one of the $\hat{\phi}$'s was unexpectedly large making a significant increase in the Monte Carlo standard error. A possible explanation for this outlier is that the computation of $\hat{\phi}$ requires the inversion of $\Delta(\hat{\mu}_j)$, $j=1, 2, \ldots, n$, and by Corollary 2.2.2 $\Delta(\hat{\mu}_j)$ may be nearly singular if one
of the elements of $\hat{\pi}_j$ gets too close to zero or if 
$1 - (\hat{\pi}_{1j} + \hat{\pi}_{2j} + \hat{\pi}_{3j})$ gets close to zero also. The statistics reported 
in Table 8.1 of Appendix A for the true probability vectors $\pi_j$'s show 
that some probabilities are very small.

The Monte Carlo properties of the estimator of $\text{Diag} \tilde{\phi}^0$ under the 
generalized multinomial variance procedure are shown in Table 5.4. The 
estimation procedure works fairly well when $\phi^0 = 1$, especially for the 
cases where $n = 100$ or $n = 300$. When $\phi^0 = 1$, $\text{Diag} \tilde{\phi}$ appears to 
approach the identity matrix as $n$ increases. When $\phi^0 = 2$ or

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$\hat{\phi}_1$</th>
<th>$\hat{\phi}_2$</th>
<th>$\hat{\phi}_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monte Carlo</td>
<td>Monte Carlo</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>n</td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>30</td>
<td>1.00</td>
<td>0.30</td>
<td>1.01</td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>0.15</td>
<td>0.99</td>
</tr>
<tr>
<td>300</td>
<td>1.00</td>
<td>0.09</td>
<td>1.00</td>
</tr>
<tr>
<td>30</td>
<td>1.81</td>
<td>0.75</td>
<td>1.80</td>
</tr>
<tr>
<td>100</td>
<td>1.88</td>
<td>0.44</td>
<td>1.90</td>
</tr>
<tr>
<td>100</td>
<td>1.91</td>
<td>0.43</td>
<td>1.96</td>
</tr>
<tr>
<td>300</td>
<td>1.99</td>
<td>0.51</td>
<td>1.99</td>
</tr>
<tr>
<td>30</td>
<td>3.36</td>
<td>1.56</td>
<td>3.33</td>
</tr>
<tr>
<td>100</td>
<td>3.71</td>
<td>1.10</td>
<td>3.71</td>
</tr>
<tr>
<td>300</td>
<td>3.87</td>
<td>0.83</td>
<td>3.90</td>
</tr>
</tbody>
</table>
110

\( \varphi^0 = 4 \), the estimation procedure, as in the case of the estimation of 
\( \varphi^0 \) with \( \hat{\varphi} \), underestimates the diagonal elements of \( \varphi^0 \). These 
bias decreases as \( n \) gets larger. The Monte Carlo standard errors for 
\( \hat{\varphi}_3 \) increases from 3.76 when \( n = 300 \), \( \varphi^0 = 4 \) to 5.58 when \( n = 300 \), 
\( \varphi^0 = 4 \). This increase is due to the maximum value taken by the \( \hat{\varphi}_3 \)'s 
which was 173.69. If this maximum is deleted the Monte Carlo standard 
deviation based on the remaining 999 estimators is 1.50. An additional 
simulation provided a Monte Carlo standard deviation of 1.54 and a Monte 
Carlo mean of 3.66.

The 5-th and 95-th percentile of the "t" statistics are presented 
on Table 5.5. These percentiles should be close to -1.64 and 1.64 which 
are the 5-th and 95-th percentiles of a standard normal distribution.

When \( \varphi^0 = 1 \) all the procedures provide percentiles close to the true 
one except for PC CARP for the case \( n = 30 \), \( m = 28 \). This could be 
because of the estimated covariance matrix (4.6.9)-(4.6.10).

When \( \varphi^0 = 2 \) or \( \varphi^0 = 4 \), it is expected that the estimated 
percentiles obtained using the logistic multinomial procedure will 
differ from the normal percentile by a multiple of \( \sqrt{\varphi^0} \). This can be 
easily seen in Table 5.5. In general, all the estimation procedures but 
the logistic multinomial one, provide estimated percentiles close to the 
normal ones, when extra variability is present. The situation improves 
as the sample size increases. It seems that in the presence of intra-
class correlation, PC CARP provides the best estimated percentiles.
Table 5.5. Estimated 5-th and 95-th percentiles\textsuperscript{a} for the "t" statistics for the coefficients estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( n )</th>
<th>( m )</th>
<th>( \phi^0 )</th>
<th>Percentile</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>5</td>
<td>-1.63</td>
<td>-1.72</td>
<td>-1.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>95</td>
<td>1.63</td>
<td>1.76</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>-1.66</td>
<td>---</td>
<td>---</td>
<td>-1.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>95</td>
<td>1.65</td>
<td>1.69</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>5</td>
<td>-1.63</td>
<td>-1.66</td>
<td>-1.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>95</td>
<td>1.66</td>
<td>1.69</td>
<td>1.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>5</td>
<td>-1.69</td>
<td>-1.71</td>
<td>-1.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>95</td>
<td>1.65</td>
<td>1.65</td>
<td>1.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>5</td>
<td>-2.23</td>
<td>-1.83</td>
<td>-1.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>95</td>
<td>2.31</td>
<td>1.86</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>-2.33</td>
<td>-1.82</td>
<td>-1.70</td>
<td>-1.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>95</td>
<td>2.33</td>
<td>1.76</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>5</td>
<td>-2.26</td>
<td>-1.66</td>
<td>-1.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>95</td>
<td>2.33</td>
<td>1.73</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>5</td>
<td>-2.35</td>
<td>-1.70</td>
<td>-1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>95</td>
<td>2.31</td>
<td>1.68</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>5</td>
<td>-3.13</td>
<td>-1.82</td>
<td>-1.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>95</td>
<td>3.11</td>
<td>1.93</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>5</td>
<td>-3.25</td>
<td>-1.75</td>
<td>-1.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>95</td>
<td>3.20</td>
<td>1.77</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>5</td>
<td>-3.21</td>
<td>-1.66</td>
<td>-1.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>95</td>
<td>3.27</td>
<td>1.72</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}The 95-th percentile of a standard normal distribution is 1.64.
Empirical biases for each of the parameters $\beta_{12}^0$, $i=1, 2, 3, \ldots, 4$ are reported on Tables 5.6-5.8 for the logistic multinomial variance and generalized multinomial variance estimation procedures. The original biases were multiplied by $10^3$ so the figures in these tables should be multiplied by $10^{-3}$. If the bias was significantly different from zero at 5%, a "*" was placed next to the

Table 5.6. Empirical bias of estimated coefficients associated with the first category

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Logistic multinomial variance</th>
<th>Generalized multinomial variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>$\beta_{12}$</td>
</tr>
<tr>
<td>n=30, m=28</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>n=100, m=28</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>n=100</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>n=300, m=28</td>
<td>1</td>
<td>-3</td>
</tr>
<tr>
<td>n=30</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>n=100</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>n=100</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>n=300</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>n=30</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>n=100</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>n=300</td>
<td>28</td>
<td>4</td>
</tr>
</tbody>
</table>

*Biases have been multiplied by $10^3$.

*Significant at 0.05.
Table 5.7. Empirical bias of estimated coefficients associated with the second category

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>φ₀</th>
<th>Procedure</th>
<th>Logistic multinomial variance</th>
<th>Generalized multinomial variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>β21</td>
<td>β22</td>
<td>β23</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>8</td>
<td>-3*</td>
<td>-2</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>74*</td>
<td>-28*</td>
<td>-34*</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>5</td>
<td>-1</td>
<td>-1*</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>-2</td>
<td>-1*</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>10</td>
<td>-4*</td>
<td>-3</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>39*</td>
<td>-13*</td>
<td>-7*</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>7</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>3</td>
<td>-0</td>
<td>-0</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>8</td>
<td>-11*</td>
<td>-11*</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>2</td>
<td>-5*</td>
<td>-4*</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>-0</td>
<td>-0</td>
<td>-0</td>
</tr>
</tbody>
</table>

*Bias have been multiplied by 10^3.

*Significant at 0.05.

reported bias. Some general conclusions can be drawn from these tables: the biases decrease as the sample size increases and the generalized multinomial variance procedure seems to provide beta estimates which have a slightly bigger bias than those estimates obtained under the logistic multinomial variance procedure. Most of the individual estimators maintain the same sign for their biases across the different combinations of n, m and φ₀. The sign of the biases also agree for the two
Table 5.8. Empirical bias of estimated coefficients associated with the third category

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Logistic multinomial variance</th>
<th>Generalized multinomial variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{31}$</td>
<td>$\beta_{32}$</td>
</tr>
<tr>
<td>n=30, m=28, $\phi^0$</td>
<td>-2</td>
<td>6*</td>
</tr>
<tr>
<td>100, 1, 1</td>
<td>-10</td>
<td>61*</td>
</tr>
<tr>
<td>100, 28, 1</td>
<td>7</td>
<td>3*</td>
</tr>
<tr>
<td>300, 28, 1</td>
<td>-3</td>
<td>1*</td>
</tr>
<tr>
<td>30, 26, 2</td>
<td>10</td>
<td>14*</td>
</tr>
<tr>
<td>100, 5, 2</td>
<td>9</td>
<td>13*</td>
</tr>
<tr>
<td>100, 26, 2</td>
<td>2</td>
<td>5*</td>
</tr>
<tr>
<td>300, 26, 2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>30, 28, 4</td>
<td>-30</td>
<td>24*</td>
</tr>
<tr>
<td>100, 28, 4</td>
<td>-1</td>
<td>5*</td>
</tr>
<tr>
<td>300, 28, 4</td>
<td>-4</td>
<td>3*</td>
</tr>
</tbody>
</table>

*Biases have been multiplied by $10^3$.

*Significant at 0.05.

estimation procedures. It seems that the presence of extra variability increases the bias under both estimation procedures.
6. BIBLIOGRAPHY


7. ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Professor Kenneth Koehler for his inspiration and guidance, his patient direction, and for all the time he spent in the preparation of this dissertation. I would also like to express my appreciation to Professor Wayne Fuller for his insights and suggestions, and for allowing me to work at his side on PC CARP.

I am grateful to the Statistics Department for giving me the opportunity to work in the Agricultural Experiment Station (AES). My three years under the supervision of Dr. David Cox, Dr. Paul Hinz and Dr. Theodore Bailey were really enjoyable. Special thanks are given to Chris Olson for efficiently typing this dissertation.

Finally, I am particularly indebted to my wife, Grecia, for her encouragement and endurance during my six years of graduate work in Ames; to my son, David, for all the hours this dissertation kept me away from him; and to my parents for all their moral support.
Table 8.1. Some Monte Carlo properties\(^a\) of the true probability vectors \(\pi_j\)'s

<table>
<thead>
<tr>
<th>Monte Carlo statistic</th>
<th>(\pi_1)</th>
<th>(\pi_2)</th>
<th>(\pi_3)</th>
<th>(1 - (\pi_1 + \pi_2 + \pi_3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.3459</td>
<td>0.3777</td>
<td>0.1785</td>
<td>0.0979</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.2672</td>
<td>0.2602</td>
<td>0.2492</td>
<td>0.1145</td>
</tr>
<tr>
<td>1st percentile</td>
<td>0.0042</td>
<td>0.0105</td>
<td>0.0001</td>
<td>0.0023</td>
</tr>
<tr>
<td>5th percentile</td>
<td>0.0189</td>
<td>0.0337</td>
<td>0.0008</td>
<td>0.0060</td>
</tr>
<tr>
<td>10th percentile</td>
<td>0.0393</td>
<td>0.0630</td>
<td>0.0020</td>
<td>0.0099</td>
</tr>
<tr>
<td>25th percentile</td>
<td>0.1101</td>
<td>0.1483</td>
<td>0.0101</td>
<td>0.0232</td>
</tr>
<tr>
<td>Median</td>
<td>0.2844</td>
<td>0.3379</td>
<td>0.0557</td>
<td>0.0558</td>
</tr>
<tr>
<td>75th percentile</td>
<td>0.5449</td>
<td>0.5739</td>
<td>0.2461</td>
<td>0.1271</td>
</tr>
<tr>
<td>90th percentile</td>
<td>0.7625</td>
<td>0.7685</td>
<td>0.6008</td>
<td>0.2437</td>
</tr>
<tr>
<td>95th percentile</td>
<td>0.8522</td>
<td>0.8520</td>
<td>0.7834</td>
<td>0.3442</td>
</tr>
<tr>
<td>99th percentile</td>
<td>0.9401</td>
<td>0.9384</td>
<td>0.9523</td>
<td>0.5506</td>
</tr>
</tbody>
</table>

\(^a\)Properties are based on 10,000 \(\pi_j\)'s.
Table 9.1. Estimated 99-th percentiles for the chi-square test of
\( H_0: \beta = \beta^0 \) when \( \phi^0 = \phi^* \)

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>( \phi^* )</th>
<th>Logistic multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>25.01</td>
<td>30.19</td>
<td>33.83</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>25.75</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>27.20</td>
<td>27.85</td>
<td>28.81</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>26.74</td>
<td>26.39</td>
<td>26.08</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>60.45</td>
<td>38.88</td>
<td>42.76</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>58.41</td>
<td>33.42</td>
<td>35.02</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>52.16</td>
<td>33.25</td>
<td>34.21</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>52.91</td>
<td>28.16</td>
<td>29.08</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>112.46</td>
<td>39.67</td>
<td>46.56</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>111.53</td>
<td>32.03</td>
<td>34.89</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>100.73</td>
<td>29.27</td>
<td>29.72</td>
</tr>
</tbody>
</table>

*The 99-th percentile of a chi-square random variable with 12 degrees of freedom is 26.22.*
Table 9.2. Estimated 95-th percentile\textsuperscript{a} for the chi-square test of 
\[ H_0: \mathbf{\theta} = \mathbf{\theta}_0 \text{ when } \mathbf{\phi}_0 = \mathbf{\phi}^* \]

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>&amp;*</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARF</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>20.97</td>
<td>23.56</td>
<td>25.34</td>
<td>21.40</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>19.61</td>
<td>---</td>
<td>---</td>
<td>22.04</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>20.98</td>
<td>22.53</td>
<td>22.70</td>
<td>25.00</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>21.08</td>
<td>21.41</td>
<td>21.66</td>
<td>22.42</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>42.79</td>
<td>28.41</td>
<td>30.82</td>
<td>26.94</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>43.61</td>
<td>25.57</td>
<td>26.77</td>
<td>25.46</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>42.70</td>
<td>23.90</td>
<td>24.29</td>
<td>25.35</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>41.67</td>
<td>21.73</td>
<td>21.93</td>
<td>22.45</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>86.98</td>
<td>30.04</td>
<td>34.47</td>
<td>27.56</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>86.71</td>
<td>25.68</td>
<td>27.68</td>
<td>28.14</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>81.80</td>
<td>23.52</td>
<td>24.02</td>
<td>23.66</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The 95-th percentile of a chi-square random variable with 12 degrees of freedom is 21.03.
Table 9.3. Estimated 90-th percentile\(^{a}\) for the chi-square test of 
\(H_0: \beta = \beta^0\) when \(\phi^0 = \phi^*\)

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>(\phi^*)</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>18.48</td>
<td>20.27</td>
<td>21.88</td>
<td>18.45</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>17.29</td>
<td>---</td>
<td>---</td>
<td>19.07</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>18.48</td>
<td>19.61</td>
<td>20.14</td>
<td>21.51</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>18.71</td>
<td>18.85</td>
<td>18.91</td>
<td>19.72</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>38.42</td>
<td>24.50</td>
<td>26.35</td>
<td>22.91</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>37.91</td>
<td>21.95</td>
<td>23.20</td>
<td>21.54</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>36.70</td>
<td>20.30</td>
<td>20.83</td>
<td>22.23</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>36.39</td>
<td>19.12</td>
<td>19.43</td>
<td>19.79</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>74.15</td>
<td>26.33</td>
<td>29.92</td>
<td>23.22</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>73.51</td>
<td>21.77</td>
<td>22.82</td>
<td>24.35</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>72.64</td>
<td>20.50</td>
<td>20.58</td>
<td>21.04</td>
</tr>
</tbody>
</table>

\(^{a}\)The 90th percentile of a chi-square random variable with 12 degrees of freedom is 18.55.
Table 9.4. Estimated 75-th percentile\textsuperscript{a} for the chi-square test of 
\( H_0: \beta = \beta^0 \) when \( \phi^0 = \phi^* \)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>n</th>
<th>m</th>
<th>( \phi^* )</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>30</td>
<td>28</td>
<td>1</td>
<td>14.78</td>
<td>15.72</td>
<td>16.70</td>
<td>14.06</td>
</tr>
<tr>
<td>PC CARP</td>
<td>30</td>
<td>28</td>
<td>1</td>
<td>13.50</td>
<td>---</td>
<td>---</td>
<td>15.25</td>
</tr>
<tr>
<td>Scaled</td>
<td>100</td>
<td>28</td>
<td>1</td>
<td>14.89</td>
<td>15.18</td>
<td>15.41</td>
<td>16.80</td>
</tr>
<tr>
<td>Generalized</td>
<td>300</td>
<td>28</td>
<td>1</td>
<td>15.19</td>
<td>15.24</td>
<td>15.20</td>
<td>16.08</td>
</tr>
<tr>
<td>Logistic</td>
<td>30</td>
<td>26</td>
<td>2</td>
<td>29.16</td>
<td>17.83</td>
<td>19.55</td>
<td>17.35</td>
</tr>
<tr>
<td>Generalized</td>
<td>100</td>
<td>5</td>
<td>2</td>
<td>29.15</td>
<td>17.11</td>
<td>17.79</td>
<td>16.90</td>
</tr>
<tr>
<td>Logistic</td>
<td>100</td>
<td>26</td>
<td>2</td>
<td>29.18</td>
<td>15.87</td>
<td>16.40</td>
<td>17.40</td>
</tr>
<tr>
<td>Generalized</td>
<td>300</td>
<td>26</td>
<td>2</td>
<td>29.31</td>
<td>15.38</td>
<td>15.68</td>
<td>15.82</td>
</tr>
<tr>
<td>Logistic</td>
<td>30</td>
<td>28</td>
<td>4</td>
<td>57.41</td>
<td>19.80</td>
<td>21.64</td>
<td>18.27</td>
</tr>
<tr>
<td>Generalized</td>
<td>100</td>
<td>28</td>
<td>4</td>
<td>57.95</td>
<td>16.96</td>
<td>17.68</td>
<td>19.03</td>
</tr>
<tr>
<td>Logistic</td>
<td>300</td>
<td>28</td>
<td>4</td>
<td>58.76</td>
<td>15.82</td>
<td>16.16</td>
<td>16.48</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The 75-th percentile of a chi-square random variable with 12 degrees of freedom is 14.85.
Table 9.5. Estimated 50-th percentile\(^a\) for the chi-square test of \(H_0: \hat{\beta} = \beta^0\) when \(\phi^0 = \phi^*\)

<table>
<thead>
<tr>
<th>(n)</th>
<th>(m)</th>
<th>(\phi^*)</th>
<th>Logistic multinomial variance</th>
<th>Scaled multinomial variance</th>
<th>Generalized multinomial variance</th>
<th>PC CARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28</td>
<td>1</td>
<td>11.25</td>
<td>11.58</td>
<td>12.18</td>
<td>10.50</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>10.54</td>
<td>---</td>
<td>---</td>
<td>11.43</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>1</td>
<td>11.37</td>
<td>11.53</td>
<td>11.62</td>
<td>12.45</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>1</td>
<td>11.47</td>
<td>11.66</td>
<td>11.79</td>
<td>12.22</td>
</tr>
<tr>
<td>30</td>
<td>26</td>
<td>2</td>
<td>21.82</td>
<td>13.18</td>
<td>14.17</td>
<td>12.79</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>2</td>
<td>22.19</td>
<td>12.34</td>
<td>13.01</td>
<td>12.66</td>
</tr>
<tr>
<td>100</td>
<td>26</td>
<td>2</td>
<td>22.54</td>
<td>12.22</td>
<td>12.63</td>
<td>13.04</td>
</tr>
<tr>
<td>300</td>
<td>26</td>
<td>2</td>
<td>22.52</td>
<td>11.66</td>
<td>11.80</td>
<td>12.16</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>4</td>
<td>41.37</td>
<td>14.33</td>
<td>15.97</td>
<td>13.11</td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>4</td>
<td>43.90</td>
<td>12.67</td>
<td>13.50</td>
<td>13.87</td>
</tr>
<tr>
<td>300</td>
<td>28</td>
<td>4</td>
<td>45.10</td>
<td>12.04</td>
<td>12.26</td>
<td>12.40</td>
</tr>
</tbody>
</table>

\(^a\)The 50-th percentile of a chi-square random variable with 12 degrees of freedom is 11.34.