EFFECTS OF MISCLASSIFICATION BIAS ON REGRESSION ANALYSES OF EPIDEMIOLOGIC DATA

by

Susan J. Reade

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 180X

September 1986
EFFECTS OF MISCLASSIFICATION BIAS ON REGRESSION ANALYSES OF EPIDEMIOLOGIC DATA

by Susan J. Reade

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1986

Approved by:

[Signatures]

Adviser

Reader

Reader
Abstract
SUSAN J. READE. Effect of Misclassification Bias on Regression Analyses of Epidemiologic Data (Under the direction of Lawrence L. Kupper.)

In most epidemiologic studies, the primary focus is on quantifying the association between exposure to some suspected agent and the development of some disease. Often, however, subjects are misclassified as to their level of exposure. For instance, in the case of a dichotomous exposure variable, a subject may be classified as unexposed when she or he is really exposed, or vice versa. Ignoring this misclassification error has been shown to introduce a bias into the estimates of certain parameters.

The research reported in the literature concerning this problem has dealt almost entirely with analyses involving dichotomous exposure variables. Little research has been done regarding situations in which there are more than two exposure categories. In addition, little research has been done regarding the effect of exposure misclassification in the presence of covariates; what has been done is limited to situations involving one dichotomous covariate. No research, thus far, has dealt with more complex (and realistic) situations involving the use of regression analyses.

For situations in which there is misclassification of exposure in a follow-up study with categorical data, we have developed a model which can be used with logistic and Poisson regression procedures. This model allows for an exposure variable with any number of categories, as well as for any number of covariates. The model helps quantify the...
biases involved when misclassification is ignored. When ancillary information is available, this model can be used to correct for the bias in the estimates produced by these regression procedures.
Acknowledgements

First and foremost, I would like to thank Larry Kupper, who has been a great teacher, advisor, and friend. It was his influence which led me to pursue a doctoral degree, and his support which helped me complete it. I not only learned an enormous amount working with him, but, more importantly, I enjoyed it as well.

I would also like to thank the other members of my committee for their comments and signatures: Drs. Gary Koch, Dave Kleinbaum, Ed Davis, and Harvey Checkoway.

There are many people who have offered me a great deal of support and encouragement throughout the writing of this dissertation. Among these are Jean Coble, whose door was always open; Elaine Walker, my roommate; Lynn Shemanski, my racquetball partner; and Kathie, Jerry, Melissa, and Christopher Edwards, my second family. Also included are Nancy Lucas, George Jerdack, Dennis Cosmatos, Jerry Schindler, and Marie Matteson, who made sure I didn’t work too hard.

Finally, I would like to thank my family, and especially my parents, David and Ruth Reade, for all the encouragement and support they have given me throughout the years.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Literature Review, Notation, and Model Formation</td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Notation</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Independence Assumption</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Proposed Model</td>
<td>7</td>
</tr>
<tr>
<td>1.5 Case-Control and Cross-Sectional Studies</td>
<td>10</td>
</tr>
<tr>
<td>1.6 Example Using Dichotomous Disease and Exposure Variables</td>
<td>13</td>
</tr>
<tr>
<td>1.6.1 Misclassification Bias for Risk Differences</td>
<td>14</td>
</tr>
<tr>
<td>1.6.2 Misclassification Bias for Risk Ratios</td>
<td>16</td>
</tr>
<tr>
<td>1.6.3 Misclassification Bias for Odds Ratios</td>
<td>17</td>
</tr>
<tr>
<td>1.7 Hypothesis Testing</td>
<td>18</td>
</tr>
<tr>
<td>1.8 Covariates</td>
<td>18</td>
</tr>
<tr>
<td>1.8.1 Covariates and Misclassification in the Literature</td>
<td>19</td>
</tr>
<tr>
<td>1.8.2 Expanded Model</td>
<td>22</td>
</tr>
<tr>
<td>1.9 Estimation of Parameters</td>
<td>24</td>
</tr>
<tr>
<td>1.10 Matrix Representation of Model (1.16)</td>
<td>26</td>
</tr>
<tr>
<td>1.11 Appropriateness of Models</td>
<td>30</td>
</tr>
</tbody>
</table>
2 Fitting and Extensions of Models

2.1 Introduction .................................................. 33
2.2 WLS and ML Estimators for Poisson and Logistic Regression Procedures .................. 34
2.3 CATMAX Computer Procedure ................................. 38
2.4 Odds Ratios and Incidence Density Ratios .................. 39
2.5 Models for Misclassification of Covariates ................. 42
  2.5.1 Development of Models .................................. 42
  2.5.2 Illustration of Model (2.7) ............................... 49
  2.5.3 Generalization of Misclassification Models .......... 52
2.6 Model to Include Interaction ................................. 53
2.7 Model to Reflect Ordinal Exposure Categories ............ 57
2.8 Nonconstant Misclassification Probabilities .............. 59

3 Properties of Estimators

3.1 Introduction .................................................. 61
3.2 Least Squares Estimation ..................................... 62
  3.2.1 Bias of $\hat{\beta}_{ls}$ and $\hat{\gamma}_{ls}$ when $\hat{\Pi}=\Pi$ .................. 63
  3.2.2 Bias of $\hat{\beta}_{ls}$ and $\hat{\gamma}_{ls}$ when $\hat{\Pi}=\Pi$ .................. 65
  3.2.3 Bias of Predicted Values ................................ 69
  3.2.4 Variance-Covariance Matrices for $\hat{\beta}_{ls}$ and $\hat{\gamma}_{ls}$ ........ 71
  3.2.5 Variance of $\hat{\Pi}\hat{\beta}_{ls}$ .............................. 73
  3.2.6 Invariance of $R^2$ and the F Statistic ................. 74
  3.2.7 Summary for LS Estimation ............................. 74
3.3 WLS Estimation .................................................. 75
  3.3.1 Bias of $\hat{\beta}_{\text{wls}}$ and $\hat{\gamma}_{\text{wls}}$ .............. 76
  3.3.2 Bias of Predicted Values .................................. 81
  3.3.3 Goodness-of-Fit Statistic for WLS Estimation .............. 82
  3.3.4 Variance-Covariance Matrices for $\hat{\beta}_{\text{wls}}$ and $\hat{\gamma}_{\text{wls}}$ 83
  3.3.5 Summary for WLS Estimation .............................. 83

3.4 ML Estimation .................................................. 84
  3.4.1 Bias of $\hat{\beta}_{\text{ml}}$ and $\hat{\gamma}_{\text{ml}}$ .............. 84
  3.4.2 Variance-Covariance Matrices for $\hat{\beta}_{\text{ml}}$ and $\hat{\gamma}_{\text{ml}}$ 87
  3.4.3 Goodness-of-Fit Statistics for ML Estimation .............. 88
  3.4.4 Summary for ML Estimation .............................. 88

4 Extension and Implications of Results

  4.1 Introduction .................................................. 90
  4.2 Misclassification of Covariates .............................. 91
  4.3 Nonconstant Misclassification Probabilities .............. 97
  4.4 Ordinal Exposure Categories ............................... 101
  4.5 Relationship Between $\hat{\beta}^*$ and $\hat{\beta}^0$ .......... 102
  4.6 Relationship Between Variances ............................ 110
  4.7 Invariance of Test of $H_0: \beta_1=\beta_2=...=\beta_k=0$ .... 113
  4.8 Relationship Between $\hat{\beta}^*$ and $\hat{\beta}$ .......... 118
  4.9 Removal of Restriction on Data for LS Estimation ....... 122
5 Numerical Applications

5.1 Introduction ............................................. 127
5.2 Conditions for Appropriateness of Models .......... 127
  5.2.1 Logistic Regression ................................. 128
  5.2.2 Poisson Regression ................................. 138
5.3 Numerical Examples .................................... 143
  5.3.1 Logistic Regression Example ...................... 146
  5.3.2 Poisson Regression Example ...................... 151
  5.3.3 Least Squares Regression Example .............. 156
5.4 Real Life Example ..................................... 160
  5.4.1 Analysis Assuming Nominal Exposure Categories 164
  5.4.2 Analysis Assuming Ordinal Exposure Categories. 170

6 Suggestions for Further Research ....................... 176

References .................................................. 178
CHAPTER ONE
LITERATURE REVIEW, NOTATION, AND MODEL FORMATION

1.1 Introduction

In most epidemiologic studies, the primary focus is on quantifying the association between exposure to some suspected agent and the development of some disease. To carry out such studies, data must be collected on subjects regarding their exposure and disease status, as well as on other risk factors for the disease under study. In some instances, obtaining the information on exposure may be a straightforward procedure. For instance, if exposure is "place of residence", acquiring the correct information would be relatively simple.

However, it is often the situation that the exposure level for an individual cannot be directly observed or measured; it may be technically infeasible or financially impractical. An example of the former situation is a cohort study examining the effect of asbestos in an occupational setting. It may be impossible to get perfect information on the level of exposure for each worker.
Instead, an estimate could be formed by piecing together relevant information. There are other techniques, for instance the use of surrogate variables, which may be employed to estimate true exposure in certain situations.

Whenever the amount of true exposure is not being directly measured, there is opportunity for individuals to be assigned an incorrect level of exposure. In the case of categorical exposure variables, this error is termed "misclassification". A subject who belongs in one exposure category may be misclassified into another. This error is often ignored in the analysis of data. Not correcting for misclassification, however, can lead to biased estimates and invalid hypothesis tests concerning the true exposure-disease relationship.

Much has been said in the literature about the bias of certain estimators in various types of studies (Bross, 1954; Diamond and Lilienfeld, 1962; Barron, 1977; Gullen et al., 1968; Goldberg, 1975; Shy et al., 1978). Most of this work, however, pertains only to situations involving dichotomous exposure and disease variables. There are few results which generalize to any number of exposure categories.

The research done thus far regarding hypothesis testing in the presence of misclassification error has also been limited. All the results relate to the use of standard chi square tests of independence. This implies that only overall tests comparing any exposure to no exposure have been examined for studies with more than two exposure levels. There are no guidelines for testing an
association between exposure and disease involving two of several exposure categories, for instance medium vs. low or high vs. low, when there is misclassification error in the study. These types of tests would be of particular importance if the researcher was interested in modeling dose-response curves.

This thesis proposes a general model which can be applied to studies with any number of exposure categories. It explains mathematically the relationship between a parameter of interest estimated from the incorrect (i.e., misclassified) data, and that parameter estimated from the (hypothetical) correct data. Using this relationship, regular least squares methods or log linear models using weighted least squares or maximum likelihood estimation can be employed to analyze data potentially involving exposure misclassification. Contrast tests can be used to test hypotheses concerning any or all of the parameters.

1.2 Notation

The theory presented in the thesis allows for situations involving two levels of disease status and any number of exposure levels or categories. Each subject in the study will have a classified exposure level and a true exposure level. These may or may not be the same.

Events of interest are:

\[ D = \text{"a subject has or develops a given disease"}\]

\[ E_i = \text{"a subject is assigned to exposure level or category } i \text{"} \]
\[ E_j^* = \text{"a subject is truly in exposure level or category } j\text{"} \]
\[ i,j = 0,1,\ldots,k \]

Throughout this paper, a "\( ^* \)" indicates the true value of a variable, its value when there is no misclassification.

With each pair \((i,j)\) of classified and true exposure categories, there are two related conditional probabilities. These will be referred to as the misclassification probabilities. One is the probability of being classified in category \(i\) given that one truly belongs in category \(j\), \(\text{pr}(E_i | E_j^*)\). For the situation in which there are just two exposure levels \([\text{exposed (subscript 1)} \text{ and unexposed (subscript 0)}]\), the two probabilities, \(\text{pr}(E_1 | E_{1^*})\) and \(\text{pr}(E_0 | E_{0^*})\) are referred to as the sensitivity and specificity, respectively.

The other conditional probability is the probability of truly belonging in category \(j\) given that one is classified in category \(i\), \(\text{pr}(E_j^* | E_i)\). Letting \(\pi_{ij} = \text{pr}(E_j^* | E_i)\), for each \(i\), we have \(\sum_{j=0}^{k} \pi_{ij} = 1\). Perfect classification will occur when \(\pi_{ij} = 0\) for all \(i \neq j\), or, equivalently, \(\pi_{ij} = 1\) when \(i=j\).

Two parameters of interest in follow-up studies are the two measures of incidence of a given disease, risk and rate. Risk is the probability of developing the disease over the time period of interest, given that one does not die from another cause first. The incidence rate is the number of new occurrences of the disease per unit of time relative to the size of the population at risk. This dissertation will primarily focus on follow-up or cohort
studies, but will also discuss case-control and cross-sectional studies.

Let $\theta_{ij} = \Pr(D \mid E_i \cap E_j^*) = \text{risk associated with having classified exposure level } i \text{ and true exposure level } j$.

$\theta_i = \Pr(D \mid E_i) = \text{risk associated with having classified exposure level } i$,

$\theta_j^* = \Pr(D \mid E_j^*) = \text{risk associated with having true exposure level } j$,

$\lambda_{ij} = \text{incidence rate associated with having classified exposure level } i \text{ and true exposure level } j$,

$\lambda_i = \text{incidence rate associated with having classified exposure level } i$,

$\lambda_j^* = \text{incidence rate associated with having true exposure level } j$.

1.3 Independence Assumption

An assumption made for the remainder of this thesis is that disease is associated only with the true exposure variable, not the classified one. In other words, given the true exposure level, the risk or rate of disease is independent of the classified exposure level.

In terms of the risk, this can be written as:

$\Pr(D \cap E_i \mid E_j^*) = \Pr(D \mid E_j^*) \Pr(E_i \mid E_j^*)$.

But,

$\Pr(D \cap E_i \mid E_j^*) = \Pr(D \mid E_i \cap E_j^*) \Pr(E_i \mid E_j^*)$. 

5
So this assumption is equivalent to
\[ \text{pr}(D|E_j^*) = \text{pr}(D|E_i \cap E_j^*), \text{ or } \quad \theta_j^* = \theta_{ij}. \] (1.1)

This says that the risk of disease for those with true exposure level \( j \) is the same for all classified exposure levels. When dealing with a follow-up study, this is a very reasonable assumption to make. Whether or not an individual develops the disease over the time period of interest should not be related to his or her classified exposure, only to the true exposure. Ignoring other risk factors, all individuals in the same true exposure category should have the same risk of disease, regardless of their classified exposure. Not only is this a reasonable assumption to make, it would be unreasonable not to make it.

Equation (1.1) can be shown to be equivalent to the condition known as "nondifferential misclassification". Nondifferential misclassification occurs, in the case of a dichotomous exposure variable, when the sensitivity and specificity are the same for the diseased and nondiseased populations. For the more general case of \((k+1)\) exposure categories, it occurs when
\[ \text{pr}(E_i | E_j^* \cap D) = \text{pr}(E_i | E_j^* \cap \bar{D}), \]
which implies
\[ \text{pr}(E_i \cap E_j^* \cap D) = \frac{\text{pr}(E_i \cap E_j^* \cap D) \text{pr}(E_j^* \cap D)}{\text{pr}(E_j^* \cap D)}. \]

Using the above expression, we can see that
\[ \text{pr}(E_i | E_j^*) = \frac{\text{pr}(E_i \cap E_j^*)}{\text{pr}(E_j^*)} = \frac{\text{pr}(E_i \cap E_j^* \cap D) + \text{pr}(E_i \cap E_j^* \cap \bar{D})}{\text{pr}(E_j^* \cap D) + \text{pr}(E_j^* \cap \bar{D})}. \]
\[
\begin{align*}
\text{pr}(E_i \cap E_j \cap \neg D) + \frac{\text{pr}(E_i \cap E_j \cdot \neg D) \cdot \text{pr}(E_j \cdot \neg D)}{\text{pr}(E_j \cdot \neg D)} \\
= \frac{\text{pr}(E_j \cdot \neg D) \cdot \left[ \text{pr}(E_j \cdot \neg D) + \text{pr}(E_j \cdot \neg D) \right]}{\text{pr}(E_j \cdot \neg D) \cdot \left[ \text{pr}(E_j \cdot \neg D) + \text{pr}(E_j \cdot \neg D) \right]} \\
= \text{pr}(E_i \mid E_j \cdot \neg D) \\
\end{align*}
\]

(1.2)

Now, \(\text{pr}(D \mid E_i \cap E_j \cdot *)\) can be written as the following:

\[
\begin{align*}
\text{pr}(D \mid E_i \cap E_j \cdot *) = \left[ \frac{\text{pr}(D \mid E_j \cdot *) \cdot \text{pr}(E_j \cdot *) \cdot \text{pr}(E_i \mid E_j \cdot \neg D)}{\text{pr}(E_i \cap E_j \cdot *)} \right] \\
= \text{pr}(D \mid E_j \cdot *) \cdot \left[ \frac{\text{pr}(E_i \mid E_j \cdot \neg D)}{\text{pr}(E_i \mid E_j \cdot *)} \right]
\end{align*}
\]

Examination of the above expression indicates that if equation (1.1) holds, then equation (1.2) will also hold, and vice versa. Therefore, the assumption of independence between the classified exposure and the risk of disease given the true exposure is equivalent to the assumption of nondifferential misclassification. Therefore, the assumption of nondifferential misclassification seems to be the only reasonable assumption to make.

1.4 Proposed Model

When a researcher is confronted with misclassification, she or he has the ability to estimate a parameter based on the
misclassified data and the desire to estimate that parameter based on the true data. For example, $\theta_i$, the risk associated with having classified exposure level $i$, can generally be estimated directly using follow-up study data, while $\theta_j^*$, the risk associated with having true exposure level $j$, cannot. Some knowledge of the misclassification probabilities involved is required to estimate $\theta_j^*$ unbiasedly.

There are various methods proposed in the literature for correcting misclassified data to obtain unbiased parameter estimates (Barron, 1977; Elton and Duffy, 1983; Copeland et. al., 1977; Shy et. al., 1978). However, these techniques are limited to studies involving dichotomous exposure and disease variables. No procedures have been suggested for studies with more than two exposure levels.

The following theory leads to a method which will give estimates of odds ratios and incidence density ratios for studies with any number of exposure categories. A mathematical relationship between the two risks described above, $\theta_i$ and $\theta_j^*$, is shown below. Use of this relationship assumes that the values of the $\pi_{ij}$'s are known, and that there is nondifferential misclassification.

$$\theta_i = \text{pr}(D | E_i) = \sum_{j=0}^{k} \text{pr}(D \cap E_j^* | E_i)$$

$$= \sum_{j=0}^{k} \frac{\text{pr}(D \cap E_j^* \cap E_i)}{\text{pr}(E_i)}$$
\[
= \sum_{j=0}^{k} \text{pr}(E_{j} \mid E_{i}) \text{pr}(D \mid E_{i} \cap E_{j})
\]

\[
= \sum_{j=0}^{k} \pi_{ij} \theta_{ij}
\]  \hspace{1cm} (1.3)

Using equation (1.1) in equation (1.3), we get

\[
\theta_{i} = \sum_{j=0}^{k} \pi_{ij} \theta_{j} \]  \hspace{1cm} (1.4)

This implies that the risk associated with a classified level is the weighted average of the risks associated with the true levels, with the weights being misclassification probabilities. This is a reasonable statement. Since \( \pi_{ij} \) is the probability that the true exposure level is \( j \) for all subjects classified into exposure level \( i \), and since \( \theta_{j} \) is the true risk for subjects in level \( j \), \( \pi_{ij} \) can be considered to be the probability that the true risk for someone classified in level \( i \) is \( \theta_{j} \). From this, expression (1.4) can be written as

\[
\theta_{i} = \sum_{j=0}^{k} \text{pr}( \text{true risk} = \theta_{j} \mid E_{i}) (\theta_{j}).
\]

Summing over the \( j \)'s in this expression is similar to the summing involved in defining a mixture of distributions.

Extending this reasoning, let \( g \) be a function of the risk or rate of disease. Also, let \( g_{i} \) be that function for subjects classified into exposure level \( i \), and \( g_{j} \) be that function for subjects truly in exposure level \( j \). Then \( \pi_{ij} \) can be considered to be the probability that the true value of the function for subjects in the \( i \)-th classified exposure level is \( g_{j} \).
As with the situation above, we can express $g_i$ in the form:

$$g_i = \sum_{j=0}^{k} \pi_{ij} g_j^*$$

(1.5)

Summing over the $g_j^*$'s takes into account all the $(k+1)$ possibilities for the true risk or rate function for subjects in the $i$-th classified exposure level. This leaves us with the weighted average, $g_i$, associated with the $i$-th classified exposure level.

There are two special cases of (1.5) which are of interest to us. First, let $g = \text{logit } \theta = \ln \left[ \theta/(1-\theta) \right]$, where $\theta$ is the risk of disease. Then we find:

$$\text{logit } \theta_i = \sum_{j=0}^{k} \pi_{ij} \text{logit } \theta_j^*$$

(1.6)

Also, let $g = \ln \lambda$, where $\lambda$ is the incidence rate. This leads to

$$\ln \lambda_i = \sum_{j=0}^{k} \pi_{ij} \ln \lambda_j^*$$

(1.7)

Expressions (1.6) and (1.7) are used later to develop regression models for risks and rates of disease.

1.5 Case-Control and Cross-Sectional Studies

Case-control studies use the probability of exposure given disease, rather than the probability of disease given exposure, to measure the association between exposure and disease. In these studies, there is often a question of correct classification of disease status as well as of exposure. An important concern is whether subjects have been properly identified as cases or controls.

Examining the effect of misclassification of disease status on
the probability of exposure given disease is analogous to examining the effect of misclassification of exposure on risk. Therefore, identity (1.4) can be used in case-control studies with nondifferential misclassification of disease simply by redefining the variables. The events of interest here are:

\[ E_1^* = "a subject is classified as, and truly is, exposed" \]
\[ D_0 = "a subject is classified as a control (not diseased)" \]
\[ D_1 = "a subject is classified as a case (diseased)" \]
\[ D_0^* = "a subject is truly a control" \]
\[ D_1^* = "a subject is truly a case" \]

Then,

\[ \text{pr}(E_1^*|D_1) = \sum_{j=0}^{1} \text{pr}(D_j^*|D_1) \text{pr}(E_1^*|D_j^*), \quad i=0,1. \]

Therefore, results which stem from model (1.4) for risk and nondifferential misclassification of exposure will also hold for this situation. For instance, it is later noted that there is bias towards the null for the risk odds ratio with nondifferential misclassification of exposure. Bias towards the null for the exposure odds ratio with nondifferential misclassification of disease is an immediate consequence of this result.

Also, if there is nondifferential misclassification of exposure in a case-control study, a relationship exists between the parameter of interest based on the misclassified data and based on the true data. Using (1.1), the relationship is the following:

\[ \text{pr}(E_1|D) = \sum_{j=0}^{k} \text{pr}(E_1 \cap E_j^*|D) \]
\[
\sum_{j=0}^{k} \frac{pr(D \mid E_i \cap E_j^*) \cdot pr(E_i \cap E_j^*)}{pr(D)} = \sum_{j=0}^{k} \frac{pr(D \mid E_j^*) \cdot pr(E_i \cap E_j^*)}{pr(D)} = \sum_{j=0}^{k} pr(E_j^* \mid D) \cdot pr(E_i \mid E_j^*)
\]

(1.8)

Notice that although this equation is similar to (1.4), the misclassification probabilities are different from those in (1.4). In the 2x2 case (k=1), these are functions of the sensitivity and specificity. Also notice that the \(pr(E_j^* \mid D)\)'s sum to 1. So, for \(pr(E_0 \mid D)\) in the 2x2 case, equation (1.8) simplifies to:

\[pr(E_0 \mid D) = pr(E_0^* \mid D) \text{ (specificity)} + [1 - pr(E_0^* \mid D)] \text{ (1-sensitivity)}\]

Many authors use these equations in some form even though they may not acknowledge them (Bross, 1954; Diamond and Lilienfeld, 1962; Keys and Kihlberg, 1963; Goldberg, 1975). The results obtained for model (1.8) are similar to the results obtained for the model (1.4) used for misclassification of exposure and measures of risk. For instance, various estimators have been found to be biased towards the null under both models. It is important to note that it is the combination of the measure of effect used (namely, the probability of disease given exposure or the probability of exposure given disease) and the misclassified variable (exposure or disease) that determines which of the models, (1.4) or (1.8), should be used in the analysis of the data.
In a cross-sectional study, the measure of effect is the prevalence of disease in the population. Since the prevalence of disease in the \( j \)-th true exposure group is \( \text{pr}(D|E_j^*) \), all calculations and conclusions found for the risk of disease also apply to the prevalence.

1.6 Example Using Dichotomous Disease and Exposure Variables

Assuming there are two levels of disease status (\( D \) and \( \bar{D} \), diseased and not diseased), and there are two levels of exposure status (\( E_1 \) and \( E_0 \), exposed and unexposed), then \( k=1 \) and

\[
E_0 = \text{"the event that a subject is assigned to the unexposed category"},
\]

\[
E_1 = \text{"the event that a subject is assigned to the exposed category"},
\]

\[
E_0^* = \text{"the event that a subject is truly in the unexposed category"},
\]

\[
E_1^* = \text{"the event that a subject is truly in the exposed category"}.
\]

The misclassification probabilities involved are:

\[
\pi_{00} = \text{pr}(E_0^* | E_0)
\]

\[
\pi_{11} = \text{pr}(E_1^* | E_1)
\]

\[
\pi_{01} = 1 - \pi_{00}
\]

\[
\pi_{10} = 1 - \pi_{11}
\]
Nondifferential misclassification implies

\[
\text{specificity} = \text{pr}(E_0 | E_0^* \cap D) = \text{pr}(E_0 | E_0^* \cap \bar{D}), \text{ and}
\]
\[
\text{sensitivity} = \text{pr}(E_1 | E_1^* \cap D) = \text{pr}(E_1 | E_1^* \cap \bar{D}).
\]

The risks involved are the following:

\[
\theta_0 = \text{risk for persons classified as unexposed}
\]
\[
\theta_1 = \text{risk for persons classified as exposed}
\]
\[
\theta_0^* = \text{risk for persons truly unexposed}
\]
\[
\theta_1^* = \text{risk for persons truly exposed}
\]

Using equation (1.4) we get:

\[
\theta_0 = \pi_{00}\theta_0^* + (1-\pi_{00})\theta_1^* = \pi_{00}(\theta_0^* - \theta_1^*) + \theta_1^* \quad (1.9)
\]
\[
\theta_1 = (1-\pi_{11})\theta_0^* + \pi_{11}\theta_1^* = \pi_{11}(\theta_1^* - \theta_0^*) + \theta_0^* \quad (1.10)
\]

1.6.1 Misclassification Bias for Risk Difference

Let \( RD = (\theta_1 - \theta_0) \) be the risk difference in the presence of misclassification and \( RD^* = (\theta_1^* - \theta_0^*) \) be the true risk difference. From equations (1.9) and (1.10):

\[
RD = \pi_{11}(\theta_1^* - \theta_0^*) + \theta_0^* + \pi_{00}(\theta_1^* - \theta_0^*) - \theta_1^*
\]
\[
= (\pi_{11} + \pi_{00} - 1)(\theta_1^* - \theta_0^*)
\]
\[
= (\pi_{11} + \pi_{00} - 1)RD^*. \quad (1.11)
\]

This is similar to a result shown by Newell (1962). His work was done in the context of case-control studies and he looked at the differences between the proportions of interest [\text{pr}(E_1 | D) and \text{pr}(E_1 | \bar{D})]. Newell noted that the true difference varied from the apparent difference by a factor of (sensitivity + specificity - 1) when there is nondifferential misclassification of exposure. This
can be derived from equation (1.8) using the same mechanics used to derive (1.11).

Equation (1.11) is the risk-based counterpart of Newell's result. With a reasonably small degree of misclassification error so that \( \frac{1}{2} < \pi_{00} < 1 \) and \( \frac{1}{2} < \pi_{11} < 1 \), equation (1.11) implies bias towards the null of the estimated risk difference. Also, if \( (\pi_{00} + \pi_{11}) < 1 \), then \( \text{RD} < 0 \) where \( \text{RD}^* > 0 \).

Newell's result implies that there is bias towards the null when estimating \( \text{pr}(E_1 \mid D) - \text{pr}(E_1 \mid \bar{D}) \), assuming the sensitivity and specificity are both greater than \( \frac{1}{2} \). Diamond and Lilienfeld (1962) present a proof in their Appendix which supports this result.

The article by Gullen et. al. (1968) considers misclassification of both exposure and disease variables and the effect of this has on the prevalence difference from a cross-sectional study. As was explained before, the theory developed for the prevalence difference also applies to the risk difference for a follow-up study.

These authors derive a formula relating the two prevalence differences involved: that from the misclassified and and that from the correctly classified population. They find that the difference involving misclassification is equal to the true difference times a factor they call \( R \). \( R \) is a function of the misclassification probabilities for exposure and disease, given in terms of \( \text{pr}(E_1 \mid E_j^*) \) and \( \text{pr}(D_1 \mid D_j^*) \).

After intense algebraic manipulations, it can be shown that

\[ R = (\text{sensitivity} + \text{specificity} \cdot -1)(\pi_{11} + \pi_{00} \cdot -1), \]
where sensitivity and specificity relate to misclassification of disease status and \( \pi_{11} \) and \( \pi_{00} \) relate to misclassification of exposure status. Thus, if there is no misclassification of exposure, this formula is identical to Newell's. If, on the other hand, there is no disease misclassification, this formula is identical to equation (1.11).

1.6.2 Misclassification Bias for Risk Ratios

As with the risk difference, it has been shown that the estimated relative risk is biased towards the null in the presence of nondifferential misclassification. Gladen and Rogan (1979) examine the effect of misclassification of exposure on the relative risk when there are any number of exposure levels. They show that the estimated relative risk comparing the highest and the lowest levels of exposure will always be biased towards the null. Below, a special case of this result is shown: when \( k=1 \), the estimated relative risk will be biased towards the null. Further, they conclude that it is impossible to make any statements regarding the direction of the bias in estimating those relative risks which compare intermediate exposure levels with the lowest.

To prove bias towards the null when \( k=1 \), let \( RR \) be the relative risk parameter estimated from the misclassified data, and let \( RR^* \) be the true relative risk. Using (1.9) and (1.10),

\[
RR = \frac{\theta_1}{\theta_0} = \frac{\pi_{11} (\theta_1^* - \theta_0^*) + \theta_0^*}{\pi_{00} (\theta_0^* - \theta_1^*) + \theta_1^*}, \quad \text{and} \quad RR^* = \frac{\theta_1^*}{\theta_0^*}
\]
Assuming $\theta_1^* > \theta_0^*$, RR must be less than RR* for there to be bias towards the null. It is possible for a switch-over to occur, i.e., for RR to be less than the null value of 1. A switch-over will only occur, however, when the degree of misclassification is very high. If $\pi_{00}$ and $\pi_{11}$ are both greater than $\frac{1}{2}$, which we will assume is the case, a switch-over is not possible.

Bias towards the null of the risk ratio, when RR*>1, is demonstrated as follows:

\[ RR < RR^* \]

\[ \frac{\pi_{11} (\theta_1^* - \theta_0^*) + \theta_0^*}{\pi_{00} (\theta_0^* - \theta_1^*) + \theta_1^*} < \frac{\theta_1^*}{\theta_0^*} \]

\[ \pi_{11} (\theta_1^* - \theta_0^*) \theta_0^* + \theta_0^* \leq \pi_{00} (\theta_0^* - \theta_1^*) \theta_1^* + \theta_1^* \]

\[ \pi_{11} \theta_0^* < -\pi_{00} \theta_1^* + (\theta_1^* + \theta_0^*) \]

\[ \pi_{11} \theta_0^* + \pi_{00} \theta_1^* < \theta_0^* + \theta_1^*, \text{ which is true since} \]

$0 < \pi_{00}, \pi_{11} < 1$. Therefore, RR is less than RR*, implying bias towards the null of the relative risk when k=1.

1.6.3 Misclassification Bias for Odds Ratios

For a case-control study with nondifferential misclassification of the exposure variable, Diamond and Lilienfeld (1962) show that the estimated exposure odds ratio is biased towards the null. Using equations (1.9) and (1.10), it can be shown that this is also the case for a risk odds ratio from a follow-up study under the same circumstances. As with the relative risk, a switch-over is not possible assuming $\pi_{00}$ and $\pi_{11}$
are both greater than \( \frac{1}{2} \).

1.7 Hypothesis Testing

Several authors have examined the effect of nondifferential misclassification error on tests of independence. Bross (1954) began by considering a test equivalent to a test of the difference between proportions in case-control situations, \( \text{pr}(E_1|D) - \text{pr}(E_1|\bar{D}) \), with misclassification of exposure. He found that the significance level is not altered by the presence of misclassification, but that the power is reduced.

The work of Gladen and Rogan (1979) deals with a test of independence of a 2xk table. The two-level factor corresponds to disease status and the k-level factor corresponds to exposure, the misclassified variable. The authors show that the Type I error rate is not affected by nondifferential misclassification but that, again, power is reduced. Mote and Anderson (1965) show this result to hold for any size contingency table.

1.8 Covariates

Invariably, there will be other risk factors for the disease, in addition to the primary exposure variable of interest, which must be considered in the analysis. Two possibilities for adjusting for these covariates are the use of stratification and the use of modeling. The literature reviewed in this section, dealing with the effects of misclassification in the presence of covariates,
relies solely on stratification as the means of adjustment. The disadvantage in the use of stratification is that, for a large number of extraneous variables and a large number of categories for each extraneous variable, the numbers in the various strata will be small, thus providing the atmosphere for imprecise results. Modeling is an alternative which is more robust to the instability problem caused by small stratum-specific sample sizes.

In the second part of this section, the theory in Section 1.4 is expanded to include covariates. Ultimately, this expanded model will be used in a regression analysis setting.

1.8.1 Covariates and Misclassification in the Literature

Greenland (1980) illustrates, by use of examples involving 2x2 tables, the effects of nondifferential misclassification of the exposure variable alone, of a confounder C alone, and of the two together. In the first of these cases, he shows that it is possible for the OR's across the strata to differ, even though the OR*’s are homogeneous. This will occur, Greenland explains, only if the covariate is associated with the classified exposure variable. It is also possible to mask heterogeneity of the OR*’s. In other words, effect modification may spuriously appear to be present or absent. This phenomenon occurs with the risk difference as well.

From results for 2x2 tables already stated, it follows that the summary OR will be biased towards the null. Greenland points out
that this is also true for the risk difference.

Greenland and Robins (1985) add that if the covariate is associated only with exposure, and not with disease nor with the rate of misclassification, then, if \((\text{sensitivity} + \text{specificity}) > 1\), adjusting for \(C\) will increase the bias due to misclassification. A proof of this is provided in their Appendix.

For a case in which a covariate is associated only with the true exposure variable, the standardized and crude OR's will be equally biased towards the null. If the covariate itself is nondifferentially misclassified, but is not a confounder, it will produce no bias in either the standardized or the crude OR.

Another situation examined by Greenland and Robins involves exposure variable misclassification rates which differ between two strata, but not between the diseased and nondiseased populations. This could occur, as is illustrated in an example, if two different methods were used to obtain information on exposure. Each method corresponds to a stratum and would presumably have a unique sensitivity and specificity. In this situation, adjustment for the covariate may or may not provide a more accurate estimate of OR than analysing the crude data. It is possible that adjusting will increase overall bias.

The authors also illustrate a result of Korn (1981) concerning hypothesis testing: if there is misclassification as described above and no other biases are involved, then under the hypothesis of no exposure-disease relationship within each stratum, the adjusted chi-square test for independence will have a
large-sample significance level no greater than the true large-sample significance level.

Finally, these results can be carried over to a regression situation. Suppose that the covariate is associated with the classified exposure variable only, and that the exposure variable is misclassified; then, since the adjusted OR is more biased than the crude, if covariates are introduced into the regression model, the estimate of OR* will be more biased than if the covariates were excluded from the model. If misclassification rates differ over the levels of the covariate, the degree of effect modification may be misrepresented. This implies that the estimated interaction coefficients may be biased.

The effects of misclassification of the confounder are also examined by Greenland. If C is a confounder such that ignoring its presence in the analysis will lead to crude odds ratios which are greater than the adjusted odds ratios obtained by controlling for C, then the stratum-specific and summary OR's will be higher than the corresponding OR*'s. If, on the other hand, ignoring C produces lower crude odds ratios than the adjusted odds ratios, then the OR's will be lower than the OR*'s. In other words, the bias will be towards the null if C has a negative confounding effect, and away from the null if C has a positive confounding effect.

It is also possible for the misclassification to result in a masking of OR* heterogeneity, or to produce a spurious or exaggerated appearance of heterogeneity. This distortion is not as
easily predicted as the one described above.

The final case involves the misclassification of both the exposure variable and the confounder. Here, Greenland assumes that the misclassifications are independent, as well as being nondifferential. Under this situation, any type of distortion is possible; there may be a bias of the stratum-specific OR's towards or away from the null, or no bias at all. As a last note, Greenland says that all of the results stated above for the effects of misclassification on the odds ratio apply to the relative risk as well.

1.8.2 Expanded Model

Suppose that the covariates of interest can be represented by a vector of p covariate values, \( \mathbf{v} = (v_1, v_2, ..., v_p) \). Also, for our purposes, we will need to assume that all covariates are categorical. Any continuous variable can be made categorical, in which case there may be some loss of information.

Each subject will fall into one of the strata defined by the combinations of categories for the grouped covariates. Each stratum will correspond to a particular set of covariate values. Suppose, for instance, that there are three categorical variables, age, sex and race, and that these variables have been categorized such that there are five age categories, two race categories, and two sex categories. Then the total number of strata will be 20, each stratum corresponding to a different age-race-sex category combination.
Let $\theta_{is}$ be the probability of disease development for a set of $N_{is}$ individuals classified into the i-th exposure category and the s-th stratum, $s=1,2,...,S$. Let $\theta_{js}^{*}$ be the probability of disease development for those individuals with true exposure category $j$, classified into the s-th stratum. Using expression (1.5) and defining $g_{is} = \logit \theta_{is}(v_s)$, where $\theta_{is}$ is as defined above and $v_s' = (v_{s1}, v_{s2},...,v_{sp})$ is the set of covariate values corresponding to the s-th stratum, then we find that

$$\logit \theta_{is}(v_s) = \sum_{j=0}^{k} \pi_{ij} \logit \theta_{js}^{*}(v_s) \quad (1.12)$$

This implies that expression (1.6) holds within each of the S strata.

It is important to notice that as we expand our model to hold within each stratum, as we are doing under model (1.12), we are inherently expanding the assumption of nondifferential misclassification to hold within each stratum. The nondifferential misclassification assumption discussed in Section 1.3 is equivalent to assuming that $\theta_{js}^{*} = \theta_{ij}$, which in turn gives

$$\theta_{1} = \sum_{j=0}^{k} \pi_{ij} \theta_{j}. \quad \text{This model form was then generalized to that of model (1.6). Therefore, if we claim that model (1.6) holds within each stratum, we are claiming that } \theta_{js}^{*} = \theta_{ij} \text{ for all } s=1,2,...,S,$$

or that nondifferential misclassification holds within each stratum.

Using similar notation, covariates can be included in model (1.7), which involves the rates of disease occurrence. Let $\lambda_{is}$ be the rate of disease development for the i-th classified exposure category and s-th stratum, and $\lambda_{js}^{*}$ be the rate of disease development for the j-th true exposure category and s-th stratum.
By using expression (1.5) and now setting $g_{is} = \ln \lambda_{is}(v_s)$, we find
\[ \ln \lambda_{is}(v_s) = \sum_{j=0}^{k} \pi_{ij} \ln \lambda_{js}^{*}(v_s) \]  
(1.13)

Again, model (1.13) implies that model (1.7) holds within each of the S strata. This requires that the nondifferential misclassification assumption holds within each stratum.

In the following section, we will show how logistic and Poisson regression methods can be used in conjunction with expressions (1.12) and (1.13), respectively, for modeling and estimation purposes.

1.9 Estimation of Parameters

Logistic regression is often used in modeling risk. It is based on fitting the logistic model
\[ \theta_{js}^{*}(v_s) = \{ 1 + \exp (-\beta_{\alpha}^{*} - \beta_j^{*} - v_s \gamma^*) \}^{-1} \]  
(1.14)

for $j=0,1,...,k$, where $\theta_{js}^{*}(v_s)$ is a probability, in this case a risk, for the $j$-th true exposure level and the stratum corresponding to the $s$-th set of $p$ covariate values $v_s = (v_{s1}, v_{s2}, ..., v_{sp})$. The vector $\gamma^* = (\gamma_1^{*}, \gamma_2^{*}, ..., \gamma_p^{*})'$ is the set of regression coefficients for the covariates. $\beta_{\alpha}^{*}$ is the reference value for the group in the lowest exposure category and first stratum. $\beta_j^{*}$ is the regression coefficient for the $j$-th true exposure level. $\beta_0^{*} = 0$, since the effect of the lowest exposure category is already included in the $\beta_{\alpha}^{*}$ term.

From (1.14) it follows that
\[ \logit \theta_{js}^{*}(v_s) = \beta_{\alpha}^{*} + \beta_j^{*} + v_s \gamma^*, \quad j=0,1,...,k \]  
(1.15)
Substituting expression (1.15) into model (1.12), we get

$$\logit \theta_{is}(v_s) = \beta_\alpha^* + \sum_{j=1}^{k} \pi_{ij} \beta_j^* + v_s'\gamma^*$$

$$= \pi_i'\beta^* + v_s'\gamma^* \quad (1.16)$$

where $\pi_i' = (1, \pi_{i1}, \ldots, \pi_{ik})$ and $\beta^* = (\beta_\alpha^*, \beta_1^*, \ldots, \beta_k^*)'$. The term $j=0$ drops out of the summation since $\beta_0^* = 0$.

Our objective is to obtain accurate (i.e., small bias and variance) estimates of the $\beta_j^*$'s. Weighted Least Squares (WLS) or Maximum Likelihood (ML) methods can be used to fit model (1.16).

In the modeling of rates, Poisson regression is frequently used. It is based on fitting the model

$$\lambda_{js^*}(v_s) = \exp (\beta_\alpha^* + \beta_j^* + v_s'\gamma^*), \quad (1.17)$$

where $\lambda_{js^*}(v_s)$ is the rate for the $j$-th true exposure level and the $s$-th stratum, and $\beta_\alpha^*, \beta_j^*, v_s, \text{and } \gamma^*$ are as defined earlier. The use of the exponential function in this model assures that estimated rates will be non-negative. From (1.17), it follows that

$$\ln \lambda_{js^*}(v_s) = \beta_\alpha^* + \beta_j^* + v_s'\gamma^*.$$ 

Substituting this expression into model (1.13) we get

$$\ln \lambda_{is}(v_s) = \beta_\alpha^* + \sum_{j=1}^{k} \pi_{ij} \beta_j^* + v_s'\gamma^*$$

$$= \pi_i'\beta^* + v_s'\gamma^* \quad (1.18)$$

Model (1.18) can be fit by WLS or ML methods to obtain an estimate of $\beta^*$.

Ordinary (unweighted) least squares methods can also be applied to a situation in which the dependent variable (say $Y$) is
continuous (e.g., blood pressure). In this case, the covariates need not be categorical variables. The subscript $s$ can refer to an individual subject rather than to a particular stratum. The same general model form would be fit, namely

$$E(Y_{is}) = \pi_i' \beta^* + v_s' \gamma^*.$$  (1.19)

1.10 Matrix Representation of Model (1.16)

In this section, model (1.16) will be expressed in matrix notation as an aid to further understanding. Model (1.16) can be written as

$$\text{logit } \theta = \Pi \beta^* + V \gamma^*$$

where

$$\begin{align*}
\text{logit } \theta &= \begin{bmatrix}
\text{logit } \theta_{01} \\
\text{logit } \theta_{11} \\
\vdots \\
\text{logit } \theta_{k1} \\
\vdots \\
\text{logit } \theta_{0S} \\
\text{logit } \theta_{1S} \\
\vdots \\
\text{logit } \theta_{kS}
\end{bmatrix} \\
\Pi &= \begin{bmatrix}
1 & \pi_{01} & \cdots & \pi_{0k} \\
1 & \pi_{11} & \cdots & \pi_{1k} \\
\vdots \\
1 & \pi_{k1} & \cdots & \pi_{kk}
\end{bmatrix}
\end{align*}$$

$$S(k+1) \times 1$$

$$S(k+1) \times (k+1)$$
\[ \beta^* = \begin{pmatrix} \beta_{a^*} \\ \beta_{1^*} \\ \vdots \\ \beta_{k^*} \end{pmatrix}_{(k+1)\times 1}, \quad V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{11} & v_{12} & \cdots & v_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{11} & v_{12} & \cdots & v_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{S_1} & v_{S_2} & \cdots & v_{Sp} \\ v_{S_1} & v_{S_2} & \cdots & v_{Sp} \\ \vdots & \vdots & \ddots & \vdots \\ v_{S_1} & v_{S_2} & \cdots & v_{Sp} \end{bmatrix}_{S(k+1)\times p}, \quad \gamma^* = \begin{pmatrix} \gamma_{1^*} \\ \gamma_{2^*} \\ \vdots \\ \gamma_{p^*} \end{pmatrix}_{p\times 1} \]

The first \((k+1)\) rows of logit \(\theta\), \(\Pi\), and \(V\) correspond to the first stratum, and the last \((k+1)\) rows to the last stratum. Notice that \(\Pi\) is a series of vertically concatenated identical matrices. Let us define the submatrix \(\Pi_1\) to be

\[ \Pi_1 = \begin{bmatrix} 1 & \pi_{01} & \cdots & \pi_{0k} \\ 1 & \pi_{11} & \cdots & \pi_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \pi_{k1} & \cdots & \pi_{kk} \end{bmatrix} \]

so that

\[ \Pi = \begin{bmatrix} \Pi_1 \\ \Pi_1 \\ \vdots \\ \Pi_1 \end{bmatrix} \]

This definition of \(\Pi\) will be used in the theory of later chapters.

Let us also define \(\Pi^0\) to be that \(\Pi\) matrix which represents perfect classification of exposure. When \(\Pi = \Pi^0\), then \(\Pi_1\) is a matrix with 1's down the first column and main diagonal and 0's elsewhere.

As an illustration, suppose there are two exposure categories (i.e., \(k=1\)) and two covariates, race and age. Further, suppose there are two levels of race and three levels of age. The number of
covariate variables needed explicitly in the model to represent race is one, the number needed to represent age is two. Therefore, there is a total of three covariate variables needed, i.e., p=3. There is a total number of six strata, i.e., S=6.

Suppose that the strata are defined as follows. The vector of covariate values for stratum s is \( \mathbf{v}_s = (v_{s1}, v_{s2}, v_{s3}) \), where

\[
\begin{align*}
  v_{s1} &= \begin{cases} 
  0, & \text{first level of race} \\
  1, & \text{second level of race} 
  \end{cases} \\
  v_{s2} &= \begin{cases} 
  1, & \text{second level of age} \\
  0, & \text{otherwise} 
  \end{cases} \\
  v_{s3} &= \begin{cases} 
  1, & \text{third level of age} \\
  0, & \text{otherwise} 
  \end{cases}
\end{align*}
\]

The parameters are interpreted as follows. \( \beta_\alpha^* \) is the reference value for the group with no exposure in the first age and first race categories. \( \beta_1^* \) is the exposure effect. \( \gamma_1^* \) is the effect for the second race category. \( \gamma_2^* \) and \( \gamma_3^* \) are the effects for the second and third age categories, respectively. Now,

\[
\logit \theta = \begin{bmatrix} 
  \logit \theta_{01} \\
  \logit \theta_{11} \\
  \logit \theta_{02} \\
  \logit \theta_{12} \\
  \logit \theta_{03} \\
  \logit \theta_{13} \\
  \logit \theta_{04} \\
  \logit \theta_{14} \\
  \logit \theta_{05} \\
  \logit \theta_{15} \\
  \logit \theta_{06} \\
  \logit \theta_{16} 
\end{bmatrix}, \\
\Pi = \begin{bmatrix} 
  1 & \pi_{01} \\
  1 & \pi_{11} \\
  1 & \pi_{01} \\
  1 & \pi_{11} \\
  1 & \pi_{01} \\
  1 & \pi_{11} \\
  1 & \pi_{01} \\
  1 & \pi_{11} \\
  1 & \pi_{01} \\
  1 & \pi_{11} \\
  1 & \pi_{01} \\
  1 & \pi_{11} 
\end{bmatrix}, \\
\beta^* = \begin{bmatrix} 
  \beta_\alpha^* \\
  \beta_1^* 
\end{bmatrix}
\]
\[ \mathbf{V} = \begin{bmatrix} v_{1}' \\
v_{1}' \\
v_{2}' \\
v_{2}' \\
v_{3}' \\
v_{3}' \\
v_{4}' \\
v_{4}' \\
v_{5}' \\
v_{5}' \\
v_{6}' \\
v_{6}' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{y}^* = \begin{bmatrix} \gamma_1^* \\
\gamma_2^* \\
\gamma_3^* \end{bmatrix} \]

\[
\text{logit } \theta = \mathbf{P}\mathbf{b}^* + \mathbf{Vy}^* \quad \text{expands to the following twelve equations:}
\]

\[
\text{logit } \theta_{01} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* \\
\text{logit } \theta_{11} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* \\
\text{logit } \theta_{02} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* + \gamma_2^* \\
\text{logit } \theta_{12} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* + \gamma_2^* \\
\text{logit } \theta_{03} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* + \gamma_3^* \\
\text{logit } \theta_{13} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* + \gamma_3^* \\
\text{logit } \theta_{04} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* + \gamma_1^* \\
\text{logit } \theta_{14} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* + \gamma_1^* \\
\text{logit } \theta_{05} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* + \gamma_1^* + \gamma_2^* \\
\text{logit } \theta_{15} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* + \gamma_1^* + \gamma_2^* \\
\text{logit } \theta_{06} = \pi_{00}\beta_{0}^* + \pi_{01}\beta_{1}^* + \gamma_1^* + \gamma_3^* \\
\text{logit } \theta_{16} = \pi_{10}\beta_{0}^* + \pi_{11}\beta_{1}^* + \gamma_1^* + \gamma_3^* 
\]

The vector of parameters \( \theta'(\theta_{01}, \theta_{11}, \theta_{02}, \theta_{12}, \ldots, \theta_{06}, \theta_{16}) \) is
estimated directly from the follow-up study data. $\hat{\theta}_{is}$ is the observed proportion of people in cell $(i,s)$ who develop the disease in question during the follow-up period. The $\pi_{ij}$'s must be specified. $\beta_{\alpha^*}$, $\beta_1^*$, $\gamma_1^*$, $\gamma_2^*$, and $\gamma_3^*$ are to be estimated using logistic regression methods.

1.11 Appropriateness of Models

Model (1.4) describes a relationship which will hold under any situation involving misclassification of exposure. However, models (1.6) and (1.7) are not models which will hold in every situation. The relationships described by these models make certain claims concerning the relationships between the distorted and true odds ratios and the distorted and true rate ratios, respectively. When the claimed relationships hold, the models are completely appropriate. When the claimed relationships hold "approximately", the models will generally perform well.

According to model (1.4), the following relationship holds in a situation involving $(k+1)$ exposure categories.

$$\theta_i - \theta_0 = \sum_{j=1}^{k} (\pi_{ij} - \pi_{0j}) (\theta_j^* - \theta_0^*)$$

or,

$$RD_i = \sum_{j=1}^{k} (\pi_{ij} - \pi_{0j}) RD_j^* , \quad \text{for } i=1,2,\ldots,k , \quad (1.20)$$

where $RD_i = (\theta_i - \theta_0)$ is the risk difference comparing the $i$-th and the lowest classified exposure categories and $RD_j^* = (\theta_j^* - \theta_0^*)$ is the risk difference comparing the $j$-th and the lowest true exposure
categories.

When \( k=1 \), expression (1.20) becomes

\[
RD = (\pi_{11} + \pi_{00} - 1) \text{RD}^*,
\]
as we saw in Section 1.6.1.

Model (1.6), which uses the parameter logit \( \theta \) rather than \( \theta \) itself, implies

\[
\logit \theta_1 - \logit \theta_0 = \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \logit \theta_j^* - \logit \theta_0^*
\]
or,

\[
\ln \text{OR}_1 = \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \ln \text{OR}_j^*, \text{ for } i=1,2,...,k, \quad (1.21)
\]

where \( \text{OR}_1 = [\theta_1(1-\theta_0)/\theta_0(1-\theta_1)] \) is the odds ratio comparing the \( i \)-th and the lowest classified exposure categories, and

\( \text{OR}_j^* = [\theta_j^*(1-\theta_0^*)/\theta_0^*(1-\theta_j^*)] \) is the odds ratio comparing the \( j \)-th and the lowest true exposure categories. Therefore, when \( k=1 \), model (1.6) implies that \( RD = (\pi_{11} + \pi_{00} - 1) \text{RD}^* \) and

\( \ln \text{OR} = (\pi_{11} + \pi_{00} - 1) \ln \text{OR}^* \), so that the same relationship which holds between the true and distorted risk differences, holds between the natural log of the true and distorted odds ratios.

Model (1.7) uses the parameter \( \ln \lambda \) rather than \( \theta \) in the model form of model (1.4). This model implies

\[
\ln \lambda_1 - \ln \lambda_0 = \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) (\ln \lambda_j^* - \ln \lambda_0^*)
\]
or,

\[
\ln \text{IDR}_1 = \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \ln \text{IDR}_j^*, \text{ for } i=1,2,...,k, \quad (1.22)
\]

where \( \text{IDR}_1 = (\lambda_1/\lambda_0) \) is the incidence density ratio comparing the \( i \)-th and the lowest classified exposure categories and
IDR_j*=(\lambda_j* / \lambda_0*) is the incidence density ratio comparing the
j-th and the lowest true exposure categories. When k=1, expression
(1.22) simplifies to
\[ \ln \text{IDR} = (\pi_{11} + \pi_{00} - 1) \ln \text{IDR}. \]

Since model (1.4) holds in any situation involving nondifferential
misclassification of exposure, the relationship described in
expression (1.20) will also hold in any such situation. If model
(1.6) is appropriate, then the relationship described in expression
(1.21) also holds. If model (1.7) is appropriate, then the
relationship described in expression (1.22) holds.

The theory developed in Chapters 2, 3, and 4 assumes that models
(1.16) and (1.18) are appropriate to a given situation. The results
from those chapters are valid as long as this is the case. Model
(1.16) will hold whenever model (1.6) holds within each stratum.
Likewise, model (1.18) will hold whenever model (1.7) holds within
each stratum. In Chapter 5, conditions are examined under which
these models will, in fact, hold well. For now, however, let us
assume that the appropriateness of these models is not an issue; let
us assume that the models fit perfectly.
CHAPTER TWO

FITTING AND EXTENSION OF MODELS

2.1 Introduction

In Chapter 1, we developed models for different types of response variables in situations involving misclassification of exposure in the presence of covariates. In this chapter, we will examine several methods for estimating the parameters in those models. Specifically, we propose the use of weighted least squares and maximum likelihood estimators based on logistic regression methods for model (1.16), Poisson regression methods for model (1.18), and regular least squares estimators for model (1.19). The formulas for the WLS and ML estimators used with logistic and Poisson regression methods are given.

Following this, the theory underlying these three models is extended to other situations involving misclassification error. These include situations involving misclassification of covariates with and without misclassification of exposure. In addition, models are developed which include interaction terms, as well as models
which can be used to fit a functional trend in the parameters.
Finally, a model is given for situations in which misclassification
probabilities differ from stratum to stratum.

2.2 WLS and ML Estimators for Poisson and Logistic
Regression Procedures

Poisson and logistic regression methods are statistical
techniques used to analyze data involving counts assumed to follow
Poisson and binomial distributions, respectively. In such cases,
the variances of the response variables will not be constant, as is
assumed in standard (unweighted) least squares regression.
Therefore, weighted least squares or maximum likelihood
estimation must be used. Weighted least squares and
maximum likelihood estimation procedures are asymptotically
equivalent. However, when cell sizes are small (e.g., <5),
maximum likelihood methods are recommended (Imrey, et. al.,

Poisson regression is used to study the relationship
between observed rates and a set of explanatory variables. One
type of Poisson regression is based on fitting log linear models
such as model (1.18). In this case, each rate is estimated by
a count (or number of incidences of disease development) divided
by the amount of population-time at risk.

Within each classified exposure level and stratum combination
(i.e., within each (1,s) cell), the observed count is assumed to be a
Poisson random variable. Further, each variable is assumed to be
independent of the others, so that the joint distribution of the data is a product Poisson likelihood. Let $X_{is}$ be the observed count (or number of occurrences) in the $i$-th classified exposure category and $s$-th stratum, and let $L_{is}$ be the total amount of population time at risk in the $i$-th classified exposure category and $s$-th stratum, $i=0,1,...,k$, $s=1,2,...,S$. Then, $X_{is}$ is assumed to follow a Poisson distribution with mean $\mu_{is} = L_{is}\lambda_{is}$. Since $\mu_{is} = X_{is}$, then $\lambda_{is} = X_{is} / L_{is}$.

The WLS estimates for model (1.18) are derived using the following formula in Grizzle, Starmer, and Koch (1969):

$$\begin{pmatrix} \hat{\beta}_{wls}^* \\ \hat{\gamma}_{wls}^* \end{pmatrix} = \left( \begin{pmatrix} \Pi' \\ V' \end{pmatrix} D_x \begin{pmatrix} \Pi & V \end{pmatrix} \right)^{-1} \begin{pmatrix} \Pi' \\ V' \end{pmatrix} D_x Y \quad (2.1)$$

where $D_x$ is a diagonal matrix of dimension $S(k+1)$, with the counts $X=(X_{10}, X_{11},...,X_{kS})'$ on the main diagonal, and $Y=(Y_{01}, Y_{11},...,Y_{kS})'$ is a $S(k+1)$ column vector with $Y_{is} = \ln(\lambda_{is})$ where $\lambda_{is} = (X_{is} / L_{is})$. $D_x^{-1}$ is the estimated covariance matrix of $Y$.

The maximum likelihood equations are based on the product Poisson likelihood for the data. Since these equations have no explicit solution, an iterative process is necessary to obtain the maximum likelihood estimates. One such process is the Newton-Raphson method. A recommended initial set of parameter estimates for this iteration procedure is the set of weighted least squares estimates given above in equation (2.1). The successive estimates are then obtained as the process changes the $t$-th step.
estimate, \( \left( \hat{\beta}_1^* \right) \), to the \((l+1)\)-th step estimate, \( \left( \hat{\beta}_{l+1}^* \right) \)

using the formula

\[
\left( \begin{array}{c}
\hat{\beta}_{l+1}^* \\
\hat{\gamma}_{l+1}^*
\end{array} \right) = \left( \begin{array}{c}
\hat{\beta}_l^* \\
\hat{\gamma}_l^*
\end{array} \right) + \left[ \begin{array}{c}
\Pi' \\
\nu'
\end{array} \right] \Sigma_l^{-1} \begin{pmatrix} \Pi & \nu \end{pmatrix} \left( X - \hat{\mu}_l \right)
\]

where \( \hat{\mu}_{1s,l} = \frac{L_{1s}}{L} \left[ \exp \left( \pi_1^r \hat{\beta}_{1s}^* + \nu_s^r \hat{\gamma}_{1s}^* \right) \right] \).

\( \hat{\mu}_l = (\hat{\mu}_{01}, \hat{\mu}_{11}, \ldots, \hat{\mu}_{lS}, \ldots, \hat{\mu}_{kS}, \ldots, \hat{\mu}_l') \) is the vector of predicted counts from the \( l \)-th estimate, and \( \Sigma_l = D \hat{\mu}_l \), where \( D \hat{\mu}_l \) is the inverse of the \( l \)-th step estimated covariance matrix of \( Y \), is a diagonal matrix with the predicted counts \( \hat{\mu}_l \) down the main diagonal. The estimate of the asymptotic covariance matrix for

\[
\left( \begin{array}{c}
\hat{\beta}_1^* \\
\hat{\gamma}_1^*
\end{array} \right)
\]

is

\[
\text{Var}_a \left( \hat{\beta}_1^*, \hat{\gamma}_1^* \right) = \left[ \begin{array}{c}
\Pi' \\
\nu'
\end{array} \right] \Sigma_l^{-1} \begin{pmatrix} \Pi & \nu \end{pmatrix}
\]

The iteration process continues until a specified "closeness" criterion is met, e.g., when the difference between two sequential sets of estimates is less than some preset value, or until a predetermined number of iterations has been performed. The final set of estimates calculated by the iteration procedure are the maximum likelihood estimates, \( \hat{\beta}_{ml}^* \) and \( \hat{\gamma}_{ml}^* \).

Logistic regression is used to analyze the relationship between an observed proportion and a set of explanatory variables. It is based on fitting linear logistic models such as model (1.16). Let \( X_{1s} \) be the number of subjects developing the
disease in the i-th classified exposure category and s-th stratum, and let \( N_{is} \) be the total number of subjects in the i-th exposure category and s-th stratum. Since each subject has a binary response (either she or he develops the disease or does not), \( X_{is} \) is assumed to follow a binomial distribution with mean \( \mu_{is} = N_{is} \theta_{is} \). The distributions for all of the exposure category and stratum combinations are assumed to be mutually independent. Therefore, the likelihood of the data is assumed to be a product binomial distribution.

The weighted least squares estimates are derived using the following formula in Grizzle, Starmer, and Koch (1969):

\[
\begin{pmatrix}
\hat{\beta}_{wls}^* \\
\hat{\gamma}_{wls}^*
\end{pmatrix}
= \left[
\begin{pmatrix}
\Pi' \\
\mathbf{V}'
\end{pmatrix}
D_{t}^{-1}
\begin{pmatrix}
\Pi \\
\mathbf{V}
\end{pmatrix}
\right]^{-1}
\begin{pmatrix}
\Pi' \\
\mathbf{V}'
\end{pmatrix}D_{t}^{-1} y
\]

where \( D_t \) is a diagonal matrix of dimension \( S(k+1) \) with the values \( t=(t_{01}, t_{11}, \ldots, t_{kS}) \) down the main diagonal where \( t_{is} = X_{is}^{-1} + (N_{is} - X_{is})^{-1} \), and \( Y=(Y_{01}, Y_{11}, \ldots, Y_{kS})' \) where \( Y_{is} = \ln \left[ \frac{X_{is}}{(N_{is} - X_{is})} \right] = \text{logit } \hat{\theta}_{is} \). \( D_t \) is the estimated covariance matrix of \( Y \).

As with the Poisson regression case, the maximum likelihood equations for this situation have no explicit solution. The Newton-Raphson iteration process used here produces the following formula for the \((l+1)\)th step of the iteration process:

\[
\begin{pmatrix}
\hat{\beta}_{l+1}^* \\
\hat{\gamma}_{l+1}^*
\end{pmatrix}
= \begin{pmatrix}
\hat{\beta}_{l}^* \\
\hat{\gamma}_{l}^*
\end{pmatrix}
+ \left[
\begin{pmatrix}
\Pi' \\
\mathbf{V}'
\end{pmatrix}
\hat{\Sigma}_l^{-1}
\begin{pmatrix}
\Pi \\
\mathbf{V}
\end{pmatrix}
\right]^{-1}
\begin{pmatrix}
\Pi' \\
\mathbf{V}'
\end{pmatrix}(p - \hat{p}_l)
\]
where $\hat{\Sigma}_l$ is the $l$-th step estimated covariance matrix of $Y, p$ is the vector of observed proportions, and $\hat{p}_l$ is the $l$-th step vector of predicted proportions. The iteration process continues until a specified criterion is met. The final set of estimates calculated by the iteration process are the maximum likelihood estimates, $\hat{\beta}_{ml}^*$ and $\hat{\gamma}_{ml}^*$.

2.3 CATMAX Computer Procedure

CATMAX is a set of statements within PROC MATRIX which is itself a part of the SAS computer package. It produces weighted least squares and maximum likelihood estimates for log-linear models, including models based on both product Poisson and binomial likelihoods. The formulas used to calculate the weighted least squares estimates are given in the previous section, as are the formulas for the maximum likelihood estimates. The iteration procedure ends when either the difference between two sequential estimates is less than .0005 or eight iterations have been performed, whichever comes first.

CATMAX provides not only the WLS and ML estimates of $\beta^*$ and $\gamma^*$, but also the vector of predicted counts, $\hat{\mu}$. Goodness-of-fit is tested for the WLS estimates by $Q_w$ (the Wald chi-square statistic), and for the ML estimates by $Q_p$ (Pearson's chi-square statistic) and by $Q_{log}$ (the log-likelihood ratio statistic). The number of degrees of freedom for each of these statistics is equal to the number of $(i,s)$ cells minus the number of parameters estimated, or $(S-1)(k+1)-p$ degrees of freedom. Also, all
address the hypothesis that the variation in the counts is compatible with the fitted model.

An option available in the CATMAX procedure is the use of contrast statements for hypothesis testing. A hypothesis of the form

$$H_0: \mathbf{C} \begin{bmatrix} \hat{\beta}^* \\ \gamma^* \end{bmatrix} = 0$$

can be tested using the Wald statistic:

$$Q_{wc} = \left( \begin{array}{c} \hat{\beta}_{ml}^* \\ \gamma_{ml}^* \end{array} \right)' \mathbf{C} \left( \mathbf{C}' \mathbf{Var}_{a}(\hat{\beta}_{ml}^*, \gamma_{ml}^*) \mathbf{C}' \right)^{-1} \mathbf{C} \left( \begin{array}{c} \hat{\beta}_{ml}^* \\ \gamma_{ml}^* \end{array} \right)$$

where $\mathbf{Var}_{a}(\hat{\beta}_{ml}^*, \gamma_{ml}^*)$ is the estimated asymptotic covariance matrix for

$$\left( \begin{array}{c} \hat{\beta}_{ml}^* \\ \gamma_{ml}^* \end{array} \right).$$

2.4 Odds Ratios and Incidence Density Ratios

When dealing with logistic regression, the parameters $\beta_1^*, \beta_2^*, ..., \beta_k^*$ are related to the true odds ratios in the following way.

Using model (1.15), we can see that

$$\ln \text{OR}_{js}^* = \logit \theta_{js}^* (v_s) - \logit \theta_{0s}^* (v_s)$$

$$= (\beta_{\alpha}^* + \beta_{j}^* + v_s' \gamma^*) - (\beta_{\alpha}^* + v_s' \gamma^*)$$

$$= \beta_{j}^*; \quad j=1,2,...,k, \ s=1,2,...,S,$$

where $\text{OR}_{js}^*$ is the odds ratio comparing the $j$-th true and the lowest true exposure categories in the $s$-th stratum. Since no interaction is assumed, $\exp(\beta_{j}^*)$ is the adjusted odds ratio comparing the $j$-th and lowest true exposure categories (i.e.,
\(\exp(\beta_j^*) = \text{OR}_j^*\). Also, the hypothesis \(H_0: \beta_j^* = 0\), for \(j = 1, 2, \ldots, k\), is equivalent to the hypothesis \(H_0: \text{OR}_j^* = 1\).

Similarly, in the case of Poisson regression, the \(\beta_j^*\)'s for \(j = 1, 2, \ldots, k\) are related to the true incidence density ratios. Using model (1.17), we can see that the following relationships hold:

\[
\ln \text{IDR}_{js}^* = \ln \lambda_{js}^*(v_s) - \ln \lambda_{0s}^*(v_s)
\]

\[
= (\beta_\alpha^* + \beta_j^* + v_s \gamma^*) - (\beta_\alpha^* + v_s \gamma^*)
\]

\[
= \beta_j^*, \quad j = 1, 2, \ldots, k, \quad s = 1, 2, \ldots, S,
\]

where \(\text{IDR}_{js}^*\) is the incidence density ratio comparing the \(j\)-th true and the lowest true exposure categories in the \(s\)-th stratum. Since no interaction is assumed, \(\exp(\beta_j^*) = \text{IDR}_{j,s}^*\) does not depend on \(s\); \(\text{IDR}_{j,s}^*\) is the adjusted incidence density ratio comparing the \(j\)-th and lowest true exposure categories.

As an illustration of the use of logistic regression with model (1.16), consider the following example.

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>Table 2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True Population</strong></td>
<td><strong>Classified Population</strong></td>
</tr>
<tr>
<td>(D)</td>
<td>(D)</td>
</tr>
<tr>
<td>(E_1^*)</td>
<td>12</td>
</tr>
<tr>
<td>(E_0^*)</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True vs. Classified Population</strong></td>
</tr>
<tr>
<td>(E_1^*)</td>
</tr>
<tr>
<td>(E_1)</td>
</tr>
<tr>
<td>(E_0)</td>
</tr>
</tbody>
</table>
Table 2.1 reflects the population when exposure is perfectly classified. Table 2.2 reflects the population when exposure is imperfectly classified with sensitivity = .9167 and specificity = .8571 in both the diseased and nondiseased populations.

Table 2.3 can be constructed using the values of the sensitivity and specificity along with the marginal totals from Table 2.1. From this table the values of the $\pi_{ij}$'s can be calculated:

\[ \pi_{00} = 120/125 = .96 \]
\[ \pi_{11} = 55/75 = .7333 \]

Note that these are the true values of the $\pi_{ij}$'s which, in this example, can be calculated from the known true population and known values of sensitivity and specificity. Normally, however, this information would not be known; the $\pi_{ij}$'s would be estimated and only the classified population, i.e., the values in Table 2.2, would be available.

Since there are no covariates in this problem, the design matrix will simply be \( \mathbf{II} = \begin{bmatrix} 1 & .04 \\ 1 & .7333 \end{bmatrix} \).

Using model (1.16) and logistic regression procedures, the following maximum likelihood estimates are calculated:

\[ \hat{\beta}_a^* = -2.18767 \]
\[ \hat{\beta}_i^* = .852971 \]

From this analysis, the estimate of the odds ratio is equal to \( \exp(\hat{\beta}_i^*) = 2.34 \), while the true odds ratio calculated from the
values in Table 2.1 is 2.25. We can see in this example that the estimate of $OR_1*$ obtained assuming model (1.6) is close to, but not equal to, $OR_1*$. This is due to the situation discussed in Section 1.11. We saw there that, while model (1.4) is always appropriate, the appropriateness of model (1.6) varies from situation to situation. Had model (1.6) been entirely appropriate here, the relationship $OR = OR_1*(\pi_{11} - \pi_{01})$ would have held exactly. Instead, $OR_1$, the odds ratio based on the classified population, is equal to 1.81 while $OR_1*(\pi_{11} - \pi_{01}) = 1.75$.

2.5 Models for Misclassification of Covariates

In this research, the models which have been developed thus far accommodate only situations for which there is just one variable misclassified, the exposure variable. Situations may arise, however, in which there is misclassification of more than just one variable. These include the possibilities of misclassification of any number of covariates along with misclassification of exposure, and the possibilities of misclassification of any number of covariates alone.

2.5.1 Development of Models

First, let us examine the problem of misclassification of covariates with no misclassification of the exposure variable in terms of model (1.16). Similar conclusions will hold for models (1.18) and (1.19). The events of interest here are:

$T_s = "a subject is assigned to stratum s"

$T_r* = "a subject is truly is stratum r"
\[ E_1^* = "a subject is assigned to, and is truly in, exposure category 1" \]

\[ D = "a subject has or develops a given disease" \]

Perfect classification of exposure implies that, if a subject is assigned to category 1, that subject is truly in category 1. Therefore, \( E_1 \), defined earlier as the event in which a subject is assigned to exposure category 1, is equivalent to the event \( E_1^* \) in which a subject is truly in exposure category 1. In a situation involving perfect classification of exposure, both \( E_1 \) and \( E_1^* \) represent the event in which a subject is assigned to, and is truly in, exposure category 1. Likewise, if there is no misclassification of covariates, \( T_s \) and \( T_s^* \) are equivalent events.

Since the strata are defined by combinations of levels of all the covariates, if one covariate is misclassified, there is a possibility for a subject to be assigned to the incorrect stratum. Incorrect stratum assignment will also occur if more than one covariate is misclassified. The possibility of a subject's being assigned to the wrong stratum is the general result when any number of covariates are misclassified. Therefore, the results reached here will hold for a situation in which there is at least one covariate misclassified and no misclassification of exposure.

Let \( \phi_{sr} = \text{pr}(T_r^* | T_s) \), \( s,r=1,2,...,S \). This is the misclassification probability which is analogous to \( \pi_{ij} \), the misclassification probability involved when there is misclassification of exposure. Now, with

\[ \theta_{isr} = \text{pr}(D | T_r^* \cap E_i^* \cap T_s) \] and \[ \phi_{sri} = \text{pr}(T_r^* | E_i^* \cap T_s) \],
we find
\[ \theta_{rs} = \text{pr}(D \mid E_i^* \cap T_s) = \sum_{r=1}^{S} \text{pr}(D \cap T_r^* \mid E_i^* \cap T_s) \]
\[ = \sum_{r=1}^{S} \text{pr}(D \mid T_r^* \cap E_i^* \cap T_s) \text{pr}(T_r^* \mid E_i^* \cap T_s) \]
\[ = \sum_{r=1}^{S} \theta_{isr} \phi_{sr} \]
\[ = \sum_{r=1}^{S} \theta_{isr} \phi_{sr} \quad (2.1) \]

assuming that \( \phi_{sr} \equiv \text{pr}(T_r^* \mid E_i^* \cap T_s) = \text{pr}(T_r^* \mid T_s) \equiv \phi_{sr} \) for all \( i=0,1,...,k \), or that the \( \phi_{sr} \)'s are constant over all exposure categories. This assumption is equivalent to an assumption concerning the independence of exposure and true covariate levels, conditional on the classified covariate levels. We can see this through the following:

\[ \text{pr}(T_r^* \mid E_i^* \cap T_s) = \text{pr}(T_r^* \mid T_s) \]

\[ \iff \frac{\text{pr}(T_r^* \cap T_s \cap E_i^*)}{\text{pr}(T_s \cap E_i^*)} = \frac{\text{pr}(T_r^* \cap T_s)}{\text{pr}(T_s)} \]

\[ \iff \frac{\text{pr}(T_r^* \cap T_s \cap E_i^*)}{\text{pr}(T_r^* \cap T_s)} = \frac{\text{pr}(T_s \cap E_i^*)}{\text{pr}(T_s)} \]

\[ \iff \text{pr}(E_i^* \mid T_r^* \cap T_s) = \text{pr}(E_i^* \mid T_s), \text{ which is equivalent to} \]

\[ \text{pr}(E_i^* \mid T_s) \text{pr}(T_r^* \mid T_s) = \text{pr}(E_i^* \mid T_r^* \cap T_s) \text{pr}(T_r^* \mid T_s) \]

\[ = \text{pr}(E_i^* \cap T_r^* \mid T_s) \]

This shows us that the assumption of constant \( \phi_{sr} \)'s over all exposure categories is equivalent to the assumption that, given
classified covariate levels, exposure and true covariate levels are independent. This may not always be the case, though. We will consider, later in this chapter, a model for situations in which the \( \phi_{sr} \)'s differ over exposure categories. For now, however, let us consider only situations in which they do not differ.

Under the situation involving misclassification of exposure, we made the assumption of nondifferential misclassification of exposure, which stated that \( \text{pr}(E_i \mid E_j^*) \) is equal for the diseased and nondiseased populations, for all \( i, j = 0, 1, \ldots, k \). In expanding our model to include covariates, the nondifferential misclassification assumption was inherently expanded to the assumption of nondifferential misclassification within each stratum, or equivalently, to the assumption that \( \text{pr}(E_i \mid E_j^* \cap T_s^*) \) is equal for the diseased and nondiseased groups for all \( i, j = 0, 1, \ldots, k \) and \( s = 1, 2, \ldots, S \).

Now we make this assumption in terms of the misclassification of the covariates, namely, we assume that \( \text{pr}(T_s \mid T_r^* \cap E_i^*) \) is equal for the diseased and nondiseased groups for all \( s, r = 1, 2, \ldots, S \) and \( i = 0, 1, \ldots, k \). This assumption can be written as:

\[
\text{pr}(T_s \mid T_r^* \cap D \cap E_i^*) = \text{pr}(T_s \mid T_r^* \cap D \cap E_i^*)
\]

which implies

\[
\text{pr}(T_s \mid T_r^* \cap D \cap E_i^*) = \text{pr}(T_s \mid T_r^* \cap E_i^*) \quad (\text{See the derivation in Section 1.3 which used } E_i^* \text{ and } E_j^* \text{ in place of } T_s \text{ and } T_r^*.)
\]

Since

\[
\theta_{isr} = \text{pr}(D \mid T_s \cap T_r^* \cap E_i^*)
\]
\[
= \text{pr}(D \mid T_r \cap E_1) \left[ \frac{\text{pr}(T_s \mid T_r \cap D \cap E_i^\star)}{\text{pr}(T_s \mid T_r \cap E_i^\star)} \right],
\]

nondifferential misclassification of the covariates within each exposure category is equivalent to

\[
\text{pr}(D \mid T_s \cap T_r^\star \cap E_i) = \text{pr}(D \mid T_r^\star \cap E_i)
\]
or \( \theta_{isr} = \theta_{ir}^\star \)

(2.2)

where \( \theta_{ir}^\star = \text{pr}(D \mid T_r^\star \cap E_i^\star) \).

Substituting expression (2.2) into model (2.1) leads to

\[
\theta_{is} = \sum_{r=1}^{S} \theta_{ir}^\star \phi_{sr}
\]

(2.3)

Model (2.3) should look familiar. It is very similar to model (1.4), developed under the situation involving misclassification of exposure. \( \theta_{ir}^\star \), the risk for exposure category \( i \) and true stratum \( r \), is analogous to \( \theta_{js}^\star \), \( \phi_{sr} \) is analogous to \( \pi_{ij} \), and summing over the \( r \)'s is analogous to summing over the \( j \)'s. In fact, the final models we will develop for misclassified covariates modify the design matrix for covariates in much the same way that the models from Chapter 1 modify the design matrix for the misclassified exposure variable.

We argued in Chapter 1 that model form (1.4) is valid not only for risks, but also for functions of risks and rates. In particular, we looked at the equality in terms of the logit of the risks and natural log of the rates. Here, the same argument can be applied to claim that model form (2.3) is valid, not only for the risks, \( \theta_{is} \) and \( \theta_{ir}^\star \), but for any function of the risks or rates,
in particular for the logit of the risks and the natural log of the rates. Looking at this model in terms of the logit of the risks we find

\[ \logit \theta_{is} = \sum_{r=1}^{S} \phi_{sr} \logit \theta_{ir^*}. \] \hspace{1cm} (2.4)

As with \( \theta_{js^*} \) in Chapter 1, we model \( \theta_{ir^*} \) using a logistic model:

\[ \theta_{ir^*} = \left\{ 1 + \exp \left( -\beta_{\alpha^*} - \beta_{i^*} - v_{r'} \gamma^* \right) \right\}^{-1} \] \hspace{1cm} (2.5)

where \( v_{r'} = (v_{r1}, v_{r2}, \ldots, v_{rp}) \) is the set of covariate values corresponding to the \( r \)-th true stratum and \( \gamma^* = (\gamma_{1^*}, \gamma_{2^*}, \ldots, \gamma_{p^*})' \) is the set of regression coefficients for the correctly classified covariates. \( \beta_{\alpha^*} \) is the predicted reference value for the group with the lowest exposure level in the first stratum. \( \beta_{i^*} \) is the incremental effect of exposure level \( i \) over the lowest exposure level. Therefore, \( \beta_0^* = 0 \) by definition.

From expression (2.5) it follows that

\[ \logit \theta_{ir^*}(v_{r'}) = \beta_{\alpha^*} + \beta_{i^*} + v_{r'} \gamma^*. \] \hspace{1cm} (2.6)

Substituting expression (2.6) into model (2.4), we find

\[ \logit \theta_{is}(v_s) = \sum_{r=1}^{S} \phi_{sr} (\beta_{\alpha^*} + \beta_{i^*} + v_{r'} \gamma^*) \]

\[ = \beta_{\alpha^*} + \beta_{i^*} + \left[ \sum_{r=1}^{S} \phi_{sr} v_{r'} \right] \gamma^* \]

\[ = \beta_{\alpha^*} + \beta_{i^*} + \phi_{s'} v_{1} \gamma^* \] \hspace{1cm} (2.7)

where \( \phi_{s'} = (\phi_{s1}, \phi_{s2}, \ldots, \phi_{sS}) \) and
\[ \mathbf{v}_1 = \begin{pmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_S \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{S1} & v_{S2} & \cdots & v_{Sp} \end{pmatrix} \]

Models similar to model (2.7) can be developed to model incidence rates and continuous response variables when there is misclassification of one or more covariates. Such a model for rates is

\[
\ln \lambda_{is}(v_s) = \beta_{a^*} + \beta_{1^*} + \phi^*_s v_1 y^* .
\] (2.8)

Likewise, the model for a continuous response variable \( y \) is

\[ \mathbb{E}(Y_{is}) = \mathbf{\beta}^*_a + \mathbf{\beta}^*_1 + \mathbf{\phi}^*_s v_1 y^* . \] (2.9)

In matrix notation, model (2.7) can be written as

\[
\logit \theta = \mathbf{\Pi}^o \mathbf{\beta}^* + \mathbf{\phi}^* v_y^* 
\]

where \( \logit \theta, \mathbf{\Pi}^o, \mathbf{\beta}^*, \mathbf{v}, \) and \( \mathbf{\gamma}^* \) are as defined in Section 1.10, and

\[
\phi = \begin{bmatrix}
\phi_{11} & 0 & \cdots & 0 & \phi_{12} & 0 & \cdots & 0 & \cdots & \phi_{1S} & 0 & \cdots & 0 \\
\phi_{11} & 0 & \cdots & 0 & \phi_{12} & 0 & \cdots & 0 & \cdots & \phi_{1S} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_{11} & 0 & \cdots & 0 & \phi_{12} & 0 & \cdots & 0 & \cdots & \phi_{1S} & 0 & \cdots & 0
\end{bmatrix} 
\]

\[
\phi = \begin{bmatrix}
\phi_{S1} & 0 & \cdots & 0 & \phi_{S2} & 0 & \cdots & 0 & \cdots & \phi_{SS} & 0 & \cdots & 0 \\
\phi_{S1} & 0 & \cdots & 0 & \phi_{S2} & 0 & \cdots & 0 & \cdots & \phi_{SS} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\phi_{S1} & 0 & \cdots & 0 & \phi_{S2} & 0 & \cdots & 0 & \cdots & \phi_{SS} & 0 & \cdots & 0
\end{bmatrix} 
\]

48
The first \((k+1)\) rows of \(\Phi\) correspond to the first stratum. These are represented in the matrix above by the top submatrix of identical rows. The last \((k+1)\) correspond to the last stratum and are represented by the lower submatrix of identical rows.

Further, as we defined \(\Pi^0\) to be the \(\Pi\) matrix which represents perfect classification of exposure, let us define \(\Phi^0\) to be that \(\Phi\) matrix which represents perfect classification of covariates. Since perfect classification of covariates implies that \(\phi_{sr}=1\) when \(s=r, s,r=1,2,\ldots,S\), it can easily be seen that \(\Phi^0 V = V\).

2.5.2 Illustration of Model (2.7)

As an illustration of the use of model (2.7), consider the following situation. Suppose there are two covariates chosen for inclusion in a model: age and sex. We could assume that either or both of the covariates are misclassified. Let us suppose that both age and sex are misclassified. Normally, of course, one would not expect a variable like sex to be misclassified.

Suppose there are three age categories and two sex categories. Define \(\psi_{ab}\) to be the probability of being classified into age category \(a\) given that one is truly in age category \(b\); \(a,b=1,2,3\). Also, define \(\xi_{cd}\) to be the probability of being classified into sex category \(c\) given that one is truly in sex category \(d\); \(c,d=1,2\). If we assume independence between the misclassification of age and sex, we find

\[
\phi_{sr} = \psi_{ab} \xi_{cd}
\]

This assumption of independence is reasonable. In general, the misclassification of one covariate should not depend on the levels of
the other covariates.

Since there are three age and two sex categories, the total number of strata is 6 (i.e., S=6). The number of covariate values used to define a stratum is 3 (i.e., p=3). Two covariate values are used to define the age category, one is used to define the sex category. Since each stratum is defined by a particular combination of age and sex categories, let us define

\[ \mathbf{v}_r' = (v_{r1}, v_{r2}, v_{r3}) = (v_{b1}, v_{b2}, v_d) \]

where \((v_{r1}, v_{r2}) = (v_{b1}, v_{b2})\) are the covariate values indicating the \(b\)-th age category; \(b=1,2,3;\)

and \((v_{r3}) = (v_d)\) is the covariate value indicating the \(d\)-th sex category; \(d=1,2.\)

Now, \(\sum_{r=1}^{S} \phi_{sr} \mathbf{v}_r'\) becomes

\[ \sum_{b=1}^{3} \sum_{d=1}^{2} \psi_{ab} \xi_{cd} (v_{b1}, v_{b2}, v_d). \]

This vector simplifies to

\[ \sum_{r=1}^{S} \phi_{sr} \mathbf{v}_r' = \left[ \sum_{b=1}^{3} \psi_{ab} (v_{b1}, v_{b2}), \sum_{d=1}^{2} \xi_{cd} (v_d) \right]. \]

Substituting the above expression into model (2.7) gives us

\[ \text{logit} \, \theta_{is}(v_s) = \beta_a \gamma^* + \beta_i \left[ \sum_{b=1}^{3} \psi_{ab} (v_{b1}, v_{b2}), \sum_{d=1}^{2} \xi_{cd} (v_d) \right] \gamma^*. \]

Suppose, now, that the following covariate values are used to
define the various categories of age and sex.

The values indicating the $b$-th age group, $(v_{b1}, v_{b2})$ for $b=1,2,3$, are:

$$(v_{11}, v_{12}) = (0,0)$$
$$(v_{21}, v_{22}) = (1,0)$$
$$(v_{31}, v_{32}) = (0,1)$$

The values indicating the $d$-th sex category, $v_d$ for $d=1,2$, are:

$v_1 = 0$
$v_2 = 1$.

Now,

$$\sum_{b=1}^{3} \psi_{ab} (v_{b1}, v_{b2})$$

$$= \psi_{a1} (0,0) + \psi_{a2} (1,0) + \psi_{a3} (0,1)$$

$$= (0,0) + \left[ \psi_{a2}, 0 \right] + \left[ 0, \psi_{a3} \right]$$

$$= \left[ \psi_{a2}, \psi_{a3} \right]$$

and

$$\sum_{d=1}^{2} \xi_{cd} v_d = \xi_{c1} (0) + \xi_{c2} (1) = \xi_{c2}$$

so that

$$\text{logit } \theta_{iS}(v_s) = \beta_{\alpha} + \beta_{\gamma} + \left[ \psi_{a2}, \psi_{a3}, \xi_{c2} \right] \gamma^*.$$  

Suppose that the stas are defined with the following covariate vectors:

$$\mathbf{v}_1' = (0,0,0)$$
$$\mathbf{v}_2' = (1,0,0)$$
$$\mathbf{v}_3' = (0,1,0)$$
$$\mathbf{v}_4' = (0,0,1)$$
$$\mathbf{v}_5' = (1,0,1)$$
$$\mathbf{v}_6' = (0,1,1).$$

The design matrix used for misclassified covariates, assuming $k=1$, 

51
is shown below. Its structure is similar to the structure of $\Pi$, the design matrix used for a misclassified exposure variable.

$$\Phi V = \begin{bmatrix}
\psi_{12} & \psi_{13} & \xi_{12} \\
\psi_{12} & \psi_{13} & \xi_{12} \\
\psi_{22} & \psi_{23} & \xi_{12} \\
\psi_{22} & \psi_{23} & \xi_{12} \\
\psi_{32} & \psi_{33} & \xi_{12} \\
\psi_{32} & \psi_{33} & \xi_{12} \\
\psi_{12} & \psi_{13} & \xi_{22} \\
\psi_{12} & \psi_{13} & \xi_{22} \\
\psi_{22} & \psi_{23} & \xi_{22} \\
\psi_{22} & \psi_{23} & \xi_{22} \\
\psi_{32} & \psi_{33} & \xi_{22} \\
\psi_{32} & \psi_{33} & \xi_{22}
\end{bmatrix}$$

and

$$\gamma^* = \begin{bmatrix}
\gamma_{1*} \\
\gamma_{2*} \\
\gamma_{3*}
\end{bmatrix}$$

The intercept parameter, $\beta^*_{\alpha}$, is the predicted reference value for the group with lowest exposure, in the first true age and sex categories. $\gamma_{1*}$ is the effect for the second true age category. $\gamma_{2*}$ is the effect for the for the third true age category. $\gamma_{3*}$ is the effect for the second true sex category.

2.5.3 Generalization of Misclassification Models

Let us reexamine models (1.16) and (2.7). Both can be written in the form

$$\text{logit } \theta = \Pi \beta^* + \Phi V \gamma^*$$

(2.11)

where the definitions of $\Pi$ and $\Phi$ depend on the misclassification
probabilities involved. Model (2.11) is equivalent to model (1.16) when $\Phi = \Phi^0$, and equivalent to model (2.7) when $\Pi = \Pi^0$. The point to be made here is that the exposure variable and the covariates are equally represented in the above model. Both are necessary to define a particular response variable.

There is nothing in the model to differentiate the roles of the exposure variable and the covariates. It is in the analysis that the two are differentiated. Therefore, it is very reasonable that the model for misclassified covariates should be a direct analogy to the model for misclassified exposure. Further, notice that the model for misclassified covariates treats the covariates as the model for misclassification of exposure treats the exposure variable. Therefore, it follows that misclassification of exposure and covariates would result in a model such as model (2.11). This model can be seen as a general model for misclassification of any number or type of variables. If there is no misclassification of either exposure or covariates, the corresponding misclassification matrices will reflect that.

2.6 Model to Include Interaction

The models developed in Chapter 1 for misclassification of exposure are strictly main effects models. They were devised ignoring any effect of interaction among the variables. If one or more of the covariates is an effect modifier, the inclusion of interaction terms in the model can be useful.

Here we will develop a model to include interaction in model
(1.16). The expanded versions of models (1.18) and (1.19) will follow directly.

Let \( w_s' = (w_{s1}, w_{s2}, \ldots, w_{sq}) \) be a set of values for possible effect modifiers for stratum \( s \) where \( q \leq p \), and let the true exposure variable be defined as an indicator variable which picks out the true exposure category. If the interaction terms are constructed by multiplying the true exposure variable by the effect modifiers, then logit \( \theta_{js}^* \) can be expressed by

\[
\text{logit } \theta_{js}^* = \beta_{j}^* + v_s' \gamma^* + \sum_{l=1}^{q} w_{sl} \delta^*_l (j-1)q+l
\]

There are \( k \) sets of \( q \) interaction terms in all, one set for each exposure effect. For \( j=0 \), the interaction terms are defined to be zero since \( \beta_0^* = 0 \). Then, the set of regression coefficients for the interaction terms is

\[
\delta^* = (\delta_1^*; \ldots, \delta_q^*; \delta_{q+1}^*; \ldots; \delta_{q+1}^*; \ldots; \delta_{k+1}^*; \ldots, \delta_{kq}^*).
\]

Summing from \( l=1 \) to \( q \) beginning with the \( ((j-1)q+1) \)th term is equivalent to summing over all the terms in the \( j \)-th set.

Now, model (1.12) becomes

\[
\text{logit } \theta_{is}(v_s) = \sum_{j=0}^{k} \pi_{ij} (\beta_{j}^* + v_s' \gamma^* + \sum_{l=1}^{q} w_{sl} \delta^*_l (j-1)q+l)
\]

\[
= \beta_{j}^* + \sum_{j=1}^{k} \pi_{ij} \beta_{j}^* + v_s' \gamma^* + \sum_{j=1}^{k} \sum_{l=1}^{q} \pi_{ij} w_{sl} \delta^*_l (j-1)q+l
\]

\[
= \beta_{j}^* + \sum_{j=1}^{k} \pi_{ij} \beta_{j}^* + v_s' \gamma^*
\]

\[
+ \sum_{l=1}^{q} \pi_{i1} w_{sl} \delta^*_l + \ldots + \sum_{l=1}^{q} \pi_{ik} w_{sl} \delta^*_l (k-1)q+l
\]
This is what we would expect, as is illustrated in the following example. Suppose there are two exposure categories and three strata. Also, let $v_s' = w_s'$ for $s=1,2,3$, i.e., all covariates are possible effect modifiers.

If $\Pi = \begin{bmatrix}
1 & \pi_{01} \\
1 & \pi_{11} \\
1 & \pi_{01} \\
1 & \pi_{11} \\
1 & \pi_{01} \\
1 & \pi_{11}
\end{bmatrix}$,

$V = \begin{bmatrix}
v_1' \\
v_1' \\
v_2' \\
v_2' \\
v_3' \\
v_3'
\end{bmatrix} = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1
\end{bmatrix}$,

$\gamma^* = \begin{bmatrix}
\gamma_1^* \\
\gamma_2^*
\end{bmatrix}$

and

$\beta^* = \begin{bmatrix}
\beta_\alpha^* \\
\beta_\delta^*
\end{bmatrix}$

then, since $v_s' = w_s'$

$w_1' = (w_{11}, w_{12}) = (v_{11}, v_{12}) = v_1' = (0, 0)$

$w_2' = (w_{21}, w_{22}) = (v_{21}, v_{22}) = v_2' = (1, 0)$

$w_3' = (w_{31}, w_{32}) = (v_{31}, v_{32}) = v_3' = (0, 1)$

and we can see that $q = 2$.

It follows that

$$\text{logit } \theta_{1s}(v_s') = \beta_\alpha^* + \sum_{j=1}^{k} \beta_j^* \pi_{ij} + v_s' \gamma^* + \sum_{l=1}^{2} \pi_{1l} w_{1l} \delta_l^*.$$  

So, for $i=0$ and $s=1$, we find

$$\sum_{l=1}^{2} \pi_{0l} w_{1l} \delta_l^* = \pi_{01} w_{11} \delta_1^* + \pi_{01} w_{12} \delta_2^* = 0;$$

for $i=0$ and $s=2$,

$$\sum_{l=1}^{2} \pi_{0l} w_{2l} \delta_l^* = \pi_{01} w_{21} \delta_1^* + \pi_{01} w_{22} \delta_2^* = \pi_{01} \delta_1^*;$$

for $i=0$ and $s=3$,

$$\sum_{l=1}^{2} \pi_{0l} w_{3l} \delta_l^* = \pi_{01} w_{31} \delta_1^* + \pi_{01} w_{32} \delta_2^* = \pi_{01} \delta_2^*;$$

for $i=1$ and $s=1,$
\[ \sum_{j=1}^{2} \pi_{11} w_{1j} \delta_{1}^{*} = \pi_{11} w_{11} \delta_{1}^{*} + \pi_{11} w_{12} \delta_{2}^{*} = 0; \]

for \( i=1 \) and \( s=2 \),
\[ \sum_{j=1}^{2} \pi_{11} w_{2j} \delta_{1}^{*} = \pi_{11} w_{21} \delta_{1}^{*} + \pi_{11} w_{22} \delta_{2}^{*} = \pi_{11} \delta_{1}^{*}; \]

for \( i=1 \) and \( s=3 \),
\[ \sum_{j=1}^{2} \pi_{11} w_{3j} \delta_{1}^{*} = \pi_{11} w_{31} \delta_{1}^{*} + \pi_{11} w_{32} \delta_{2}^{*} = \pi_{11} \delta_{2}^{*}; \]
so that

\[
\text{logit } \theta_{01} = \beta_{\alpha}^{*} + \pi_{01} \beta_{1}^{*},
\]

\[
\text{logit } \theta_{11} = \beta_{\alpha}^{*} + \pi_{11} \beta_{1}^{*},
\]

\[
\text{logit } \theta_{02} = \beta_{\alpha}^{*} + \pi_{01} \beta_{1}^{*} + \gamma_{1}^{*} + \pi_{01} \delta_{1}^{*},
\]

\[
\text{logit } \theta_{12} = \beta_{\alpha}^{*} + \pi_{11} \beta_{1}^{*} + \gamma_{1}^{*} + \pi_{11} \delta_{1}^{*},
\]

\[
\text{logit } \theta_{03} = \beta_{\alpha}^{*} + \pi_{01} \beta_{1}^{*} + \gamma_{2}^{*} + \pi_{01} \delta_{2}^{*},
\]

\[
\text{and logit } \theta_{13} = \beta_{\alpha}^{*} + \pi_{11} \beta_{1}^{*} + \gamma_{2}^{*} + \pi_{11} \delta_{2}^{*}.
\]

This implies that the model including interaction can be written in matrix notation as

\[ \text{logit } \theta = \Pi \beta^{*} + \mathbf{V}\gamma^{*} + \mathbf{W}\delta^{*} \]

where

\[
\mathbf{W} = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
\pi_{01} & 0 \\
\pi_{11} & 0 \\
0 & \pi_{01} \\
0 & \pi_{11}
\end{bmatrix} \quad \text{and} \quad \delta^{*} = \begin{bmatrix}
\delta_{1}^{*} \\
\delta_{2}^{*}
\end{bmatrix}.
\]

56
2.7 Model to Reflect Ordinal Exposure Categories

Let us return to the situation of misclassification of exposure alone. Often, categories of the exposure variable are ordinal rather than nominal. If so, it may be of interest to fit some functional trend to the exposure parameters. Here, we will develop a model which accommodates such situations.

We will begin by developing a model to fit a linear trend in the true exposure effects. A linear trend implies that the parameters follow the functional form: \( \beta_j^* = \beta_\alpha^* + c_j \tau^* \), where \( c_j \) is a score assigned to true exposure category \( j \), for \( j=0,1,...,k \). A simple example is \( c_j=j \) which assumes equal spacing of the exposure levels. The interpretation of the parameters here is different from the case in which the exposure categories are nominal. \( \beta_\alpha^* \) is the intercept term; and \( \beta_0^* = \beta_\alpha^* + c_0 \tau^* \), rather than zero as it was for nominal exposure categories. Finally, \( \tau^* \) is the linear effect for exposure.

The vector of exposure parameters is now \( \beta^* = \begin{pmatrix} \beta_\alpha^* \\ \tau^* \end{pmatrix} \).

The logistic model becomes

\[
\theta_{js^*}(v_s) = \left\{ 1 + \exp(-\beta_\alpha^* - c_j \tau^* - v_s' \gamma^*) \right\}^{-1},
\]

so that

\[
\logit \theta_{js^*}(v_s) = \beta_\alpha^* + c_j \tau^* + v_s' \gamma^*.
\]

Using this expression in model (1.12) gives us

\[
\logit \theta_{is}(v_s) = \sum_{j=0}^{k} \pi_{ij} (\beta_\alpha^* + c_j \tau^* + v_s' \gamma^*)
\]

\[
= \beta_\alpha^* + \tau^* \sum_{j=0}^{k} \pi_{ij} c_j + v_s' \gamma^* \quad \text{(2.10)}
\]
In general, suppose we want to fit the following functional trend in the exposure parameters:

\[ \beta_j = \beta_{\alpha} + c_j \tau_1 + c_j \tau_2 + \ldots + c_j \tau_m \]

where \( m < k \). A quadratic trend would be fitted when \( m = 2 \), a cubic when \( m = 3 \), and so on. The vector of exposure parameters is

\[ \beta^* = \begin{pmatrix} \beta_{\alpha}^* \\ \tau_1^* \\ \tau_2^* \\ \vdots \\ \tau_m^* \end{pmatrix} \]

In this case,

\[ \logit \theta_{js}^*(\nu_s) = \beta_{\alpha}^* + c_j \tau_1^* + c_j \tau_2^* + \ldots + c_j \tau_m^* + \nu_s \gamma^* \]

and, when substituting this expression into model (1.12), we find

\[ \logit \theta_{js}^*(\nu_s) = \beta_{\alpha}^* + \tau_1^* \sum_{j=0}^{k} c_j \pi_1j + \tau_2^* \sum_{j=0}^{k} c_j \pi_2j + \ldots + \tau_m^* \sum_{j=0}^{k} c_j \pi_mj + \nu_s \gamma^* \] (2.11)

As an illustration of this model, consider a situation in which the objective is to fit a linear trend to data for which there are four exposure categories (i.e., \( k = 3 \)) and \( c_j = j, j=0,1,2,3 \). It follows from model (2.10) that the design matrix would take the form \( \Pi = \begin{bmatrix} \Pi_1 \\ \Pi_1 \\ \vdots \\ \Pi_1 \end{bmatrix} \) where

58
\[ \Pi_i = \begin{bmatrix} 1 & \pi_{01} + 2\pi_{02} + 3\pi_{03} \\ 1 & \pi_{11} + 2\pi_{12} + 3\pi_{13} \\ 1 & \pi_{21} + 2\pi_{22} + 3\pi_{23} \\ 1 & \pi_{31} + 2\pi_{32} + 3\pi_{33} \end{bmatrix} \quad \text{and} \quad \beta^* = \begin{bmatrix} \beta^* \\ \tau^* \end{bmatrix}. \]

Notice that, in this case, the elements of the \( \Pi \) matrix are not probabilities. Instead, the elements of the second column are weighted averages of the scores, \( c_j = j, \ j = 0, 1, 2, 3 \), with the weights being the misclassification probabilities.

2.8 Nonconstant Misclassification Probabilities

Throughout this and the previous chapters, we have made certain restrictions on the misclassification probabilities. In the development of models (1.16), (1.18), and (1.19) we assumed that the \( \pi_{ij} \)'s do not differ from stratum to stratum. Similarly, in the development of models (2.7), (2.8) and (2.9), we assumed that the \( \phi_{sr} \)'s do not differ over exposure categories.

These assumptions may not always be true. Suppose, for instance, that the methods for classifying subjects into exposure categories varied from city to city. A covariate which represents "city" included in the model will lead to different \( \phi_{sr} \)'s over the strata.

Modifying the models to accommodate situations such as the one described above is a very simple matter. For instance, model (1.16) becomes

\[
\text{logit } \theta_{is}(v_s) = \beta^{*\alpha} + \sum_{j=1}^{k} \beta^{*\pi_{ij}s} + v_s' \gamma^* 
\]
where \( \pi_{ij} = \text{pr}(E_{j}^{*} \mid E_{i} \cap T_{s}) \) so that each stratum has a distinct set of \( \pi_{ij} \)'s associated with it. Models (1.18) and (1.19) are modified similarly.

Likewise, if the \( \phi_{sr} \)'s differ over exposure categories, model (2.7) becomes

\[
\text{logit } \theta_{is}(v_{s}) = \beta_{\alpha}^{*} + \beta_{\iota}^{*} + \sum_{r=1}^{S} \phi_{sri} v_{r}^{*} y_{s}^{*},
\]

where \( \phi_{sri} = \text{pr}(T_{r}^{*} \mid T_{s} \cap E_{i}^{*}) \) so that each exposure category has a distinct set of \( \phi_{sr} \)'s associated with it. Models (2.8) and (2.9) are modified similarly.
CHAPTER THREE

PROPERTIES OF ESTIMATORS

3.1 Introduction

In Chapters 1 and 2, we developed models to accommodate several different types of situations involving misclassification error. In Chapter 2, we also considered different methods for estimating parameters associated with these models. Here, we will examine the properties, i.e., bias and variance, of these different types of estimators. We will be dealing specifically with those models derived in Chapter 1 for misclassification of exposure, namely models (1.16), (1.18), and (1.19). One general assumption made in this and in the subsequent chapter is that models (1.16), (1.18), and (1.19) are appropriate. That is, we are assuming that the conditions required for models (1.16) and (1.18) to be valid (see Chapter 1) hold.

One important consideration in examining the properties of the estimators is the following. Models (1.16), (1.18), and (1.19) were derived assuming that the II matrix was known. However, this is usually not the case; instead, the values of
the elements in \( \Pi \) are estimated (or guessed at) by the investigator. We are interested in the properties of certain estimators when an estimate of \( \Pi \), rather than \( \Pi \) itself, is used in the fitted model. Define the matrix \( \hat{\Pi} \) to be an estimate of the matrix \( \Pi \). In addition, define \( \hat{\Pi}^0 \) to be that estimate of \( \Pi \) which reflects no misclassification of exposure. In the following sections we will consider a fitted model which uses \( \hat{\Pi} \) rather than \( \Pi \), and the special case of \( \hat{\Pi} = \hat{\Pi}^0 \) to determine the consequences of ignoring misclassification error.

Let us begin by examining the LS estimators of the parameters in model (1.19). The primary reason for examining the least squares estimators first is that they are the least complex and most commonly used. The WLS and ML estimators will be discussed later in a similar fashion.

3.2 Least Squares Estimation

Let \( Y = (Y_{01}, Y_{11}, \ldots, Y_{kS})' \) be a vector of continuous response variables. The covariates in this situation need not be categorical. If they are, instead, continuous, \( Y_{is} \) refers to the \( s \)-th subject among those in the \( i \)-th classified exposure category. The \( \pi_{ij} \)'s are assumed to be the same for all subjects; the misclassification probabilities cannot vary from subject to subject within a classified exposure category.

Further, it is assumed that each classified exposure category contains the same number of subjects, \( S \). This may not be a realistic requirement to put on the data. For now, however, we will
assume that it holds. In Chapter 4 we will investigate the situation in which this requirement is violated.

As was discussed in the previous section, the estimate of \( \Pi \) used in fitting the model may not always be the true \( \Pi \) matrix. Therefore, we are faced with a situation involving two models:

the true model, \( E(Y) = \Pi \beta^* + \nu Y^* \)  \hspace{1cm} (3.1)

and the model to be fitted, \( E(Y) = \hat{\Pi} \beta + \nu Y \)  \hspace{1cm} (3.2)

where \( \beta = (\beta_\alpha, \beta_1, \ldots, \beta_k)' \) is the set of regression coefficients for the exposure variables, and \( Y = (Y_1, Y_2, \ldots, Y_p)' \) is the set of regression coefficients for the covariates.

3.2.1 Bias Of \( \hat{\beta}_{ls} \) And \( \hat{\gamma}_{ls} \) When \( \hat{\Pi} = \Pi \)

The following derivation (Seber, 1976) gives formulas for \( \hat{\beta}_{ls} \) and \( \hat{\gamma}_{ls} \), the LS estimators of \( \beta \) and \( \gamma \), from which we can investigate the properties of the two. Model (3.2) can be written

\[
Y = \hat{\Pi} \beta + \nu Y + e
\]

where \( e = (e_{01}, e_{11}, \ldots, e_{kS})' \) is a vector of error terms. Now we compute the LS estimates as follows:

\[
e'e = (Y - \hat{\Pi} \beta - \nu Y)'(Y - \hat{\Pi} \beta - \nu Y) = Y'Y - 2\beta'\hat{\Pi}'\nu Y - 2\gamma'Y + 2\beta'\hat{\Pi}'\nu Y + \beta'\hat{\Pi}'\nu Y + \gamma'\nu'Y
\]

\[
\Rightarrow \frac{\partial (e'e)}{\partial \beta} = -2\hat{\Pi}'Y + 2\hat{\Pi}'\nu Y + 2\hat{\Pi}'\nu Y + \hat{\Pi}' \beta
\]  \hspace{1cm} (3.3)

\[
\Rightarrow \frac{\partial (e'e)}{\partial \gamma} = -2\nu'Y + 2\nu'\hat{\Pi} \beta + 2\nu'Y
\]  \hspace{1cm} (3.4)

Setting expression (3.3) equal to 0 implies

\[
\hat{\beta}_{ls} = (\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi}' (Y - \nu \hat{\gamma}_{ls})
\]  \hspace{1cm} (3.5)
Setting expression (3.4) equal to 0 implies, along with expression (3.5), that
\[ \mathbf{V}'\mathbf{V}_{\hat{\gamma}_{ls}} = \mathbf{V}'\mathbf{Y} - \mathbf{V}' \left[ \mathbf{\hat{\Pi}} \left( \mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}} \right)^{-1} \mathbf{\hat{\Pi}}' \left( \mathbf{\hat{Y}} - \mathbf{V}_{\hat{\gamma}_{ls}} \right) \right] \]
\[ \Rightarrow \mathbf{V}' \left[ \mathbf{I}_{S(k+1)} - \mathbf{\hat{\Pi}} \left( \mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}} \right)^{-1} \mathbf{\hat{\Pi}}' \right] \mathbf{V}_{\hat{\gamma}_{ls}} \]
\[ = \mathbf{V}' \left[ \mathbf{I}_{S(k+1)} - \mathbf{\hat{\Pi}} \left( \mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}} \right)^{-1} \mathbf{\hat{\Pi}}' \right] \mathbf{\hat{Y}} \]
\[ \Rightarrow \mathbf{V}'\mathbf{R}\mathbf{V}_{\hat{\gamma}_{ls}} = \mathbf{V}'\mathbf{\hat{Y}} \]
\[ \Rightarrow \mathbf{\hat{\gamma}_{ls}} = (\mathbf{V}'\mathbf{R}\mathbf{V})^{-1} \mathbf{V}'\mathbf{\hat{Y}} \quad (3.6) \]

where \( \mathbf{R} = \left[ \mathbf{I}_{S(k+1)} - \mathbf{\hat{\Pi}} \left( \mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}} \right)^{-1} \mathbf{\hat{\Pi}}' \right] \) and \( \mathbf{I}_{S(k+1)} \) is an identity matrix of dimension \( S(k+1) \).

Now we can examine the bias of \( \hat{\mathbf{\beta}}_{ls} \) and \( \hat{\mathbf{\gamma}}_{ls} \) by conditioning on \( \mathbf{\hat{\Pi}} \), which is equivalent to assuming that \( \mathbf{\hat{\Pi}} \) is a matrix of known constants. The expected value of \( \hat{\mathbf{\beta}}_{ls} \) given \( \mathbf{\hat{\Pi}} \) is
\[
\mathbb{E}(\hat{\mathbf{\beta}}_{ls} | \mathbf{\hat{\Pi}}) = (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \left( \mathbf{\Pi}\mathbf{\beta}^* + \mathbf{V}\mathbf{\gamma}^* \right)
- (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \mathbf{V}' (\mathbf{V}'\mathbf{R}\mathbf{V})^{-1} \mathbf{V}' \mathbf{\hat{R}} (\mathbf{\Pi}\mathbf{\beta}^* + \mathbf{V}\mathbf{\gamma}^*)
= (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\Pi}\mathbf{\beta}^* + (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \mathbf{V}\mathbf{\gamma}^*
- (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \mathbf{V}' (\mathbf{V}'\mathbf{R}\mathbf{V})^{-1} \mathbf{V}' \mathbf{\hat{R}} \mathbf{\Pi}\mathbf{\beta}^*
- (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \mathbf{V}\mathbf{\gamma}^*$
\[
= (\mathbf{\hat{\Pi}}' \mathbf{\hat{\Pi}})^{-1} \mathbf{\hat{\Pi}}' \left[ \mathbf{I}_{S(k+1)} - \mathbf{V}' (\mathbf{V}'\mathbf{R}\mathbf{V})^{-1} \mathbf{V}' \mathbf{\hat{R}} \right] \mathbf{\Pi}\mathbf{\beta}^* \quad (3.7)
\]

Notice that when \( \mathbf{\hat{\Pi}} = \mathbf{\Pi} \), \( \mathbf{\hat{\Pi}}\mathbf{\Pi} = \mathbf{0} \) where \( \mathbf{0} \) is a matrix whose elements are all zeros. It follows that \( \mathbb{E}(\hat{\mathbf{\beta}}_{ls} | \mathbf{\hat{\Pi}} = \mathbf{\Pi}) = \mathbf{\beta}^* \).

Therefore, if the correct \( \mathbf{\hat{\Pi}} \) matrix is used in the fitted model (i.e., \( \mathbf{\hat{\Pi}} = \mathbf{\Pi} \)), \( \hat{\mathbf{\beta}}_{ls} \) is an unbiased estimator of \( \mathbf{\beta}^* \).

The expected value of \( \hat{\mathbf{\gamma}}_{ls} \) given \( \mathbf{\hat{\Pi}} \) is
\[ E(\hat{\gamma}_{1s} | \hat{\Pi}) = (V^T R V)^{-1} V^T R (\Pi \beta^* + \nu \gamma^*) \]
\[ = (V^T R V)^{-1} V^T R \Pi \beta^* + (V^T R V)^{-1} V^T R \nu \gamma^* \]
\[ = (V^T R V)^{-1} V^T R \Pi \beta^* + \gamma^* \] (3.8)

Again, if \( \hat{\Pi} = \Pi \), \( \hat{R} \Pi = 0 \) and \( E(\hat{\gamma}_{1s} | \hat{\Pi} = \Pi) = \gamma^* \) so that \( \hat{\gamma}_{1s} \) is an unbiased estimator of \( \gamma^* \) when \( \hat{\Pi} = \Pi \).

3.2.2 Bias Of \( \hat{\beta}_{1s} \) And \( \hat{\gamma}_{1s} \) When \( \hat{\Pi} \neq \Pi \)

By investigating the matrix \( \hat{R} \), we can gain some insight regarding the bias of \( \hat{\beta}_{1s} \) and \( \hat{\gamma}_{1s} \) when \( \hat{\Pi} \) is not equal to \( \Pi \).

Expand the quantity \( \hat{\Pi}(\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi} \) as follows:

Let \( \hat{\Pi} = \begin{bmatrix} \hat{\Pi}_1' \\ \hat{\Pi}_1 \\ \vdots \\ \hat{\Pi}_1 \end{bmatrix} \) where \( \hat{\Pi}_1 \) is the estimate of \( \Pi_1 \) and has dimensions \( (k+1) \times (k+1) \);

then,

\[ (\hat{\Pi}' \hat{\Pi}) = \begin{bmatrix} \hat{\Pi}_1', \hat{\Pi}_1', ..., \hat{\Pi}_1' \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_1 \\ \vdots \\ \hat{\Pi}_1 \end{bmatrix} = S (\hat{\Pi}_1' \hat{\Pi}_1) \]

\[ \Rightarrow (\hat{\Pi}' \hat{\Pi})^{-1} = \frac{1}{S} (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \]
\[ \Rightarrow (\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi}' = \frac{1}{S} (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \begin{bmatrix} \hat{\Pi}_1', \hat{\Pi}_1', ..., \hat{\Pi}_1' \end{bmatrix} \]
\[ = \frac{1}{S} \begin{bmatrix} (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1', (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1', ..., (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1' \end{bmatrix} \] (3.9)
\[ \Rightarrow \hat{\Pi} (\hat{\Pi}' \hat{\Pi}) \hat{\Pi}' = \]
\[
\frac{1}{S} \left[ \hat{\Pi}_1 \right] \left[ \begin{array}{c} \hat{\Pi}_1^{-1} \hat{\Pi}_1' \hat{\Pi}_1^{-1} \hat{\Pi}_1' \ldots \hat{\Pi}_1^{-1} \hat{\Pi}_1' \\ \hat{\Pi}_1^{-1} \hat{\Pi}_1' \hat{\Pi}_1^{-1} \hat{\Pi}_1' \ldots \hat{\Pi}_1^{-1} \hat{\Pi}_1' \\ \vdots \ \vdots \ \vdots \\ \hat{\Pi}_1^{-1} \hat{\Pi}_1' \hat{\Pi}_1^{-1} \hat{\Pi}_1' \ldots \hat{\Pi}_1^{-1} \hat{\Pi}_1' \end{array} \right] 
\]

Recall that \( \hat{\Pi}_1 \) is a square matrix. It has no linear dependencies among the rows or columns and is therefore invertible. This implies that

\[
\hat{\Pi}_1 (\hat{\Pi}_1' \hat{\Pi}_1^{-1}) \hat{\Pi}_1' = \hat{\Pi}_1 \hat{\Pi}_1^{-1} \hat{\Pi}_1' = \mathbf{I}_{k+1}
\]

so that \( \hat{\Pi} (\hat{\Pi}' \hat{\Pi}^{-1}) \hat{\Pi}' = \frac{1}{S} \left[ \begin{array}{cccc} I_{k+1} & I_{k+1} & \cdots & I_{k+1} \\ I_{k+1} & I_{k+1} & \cdots & I_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ I_{k+1} & I_{k+1} & \cdots & I_{k+1} \end{array} \right] \) (3.10)

This implies that \( \hat{R} = [ \mathbf{I}_{S(k+1)} - \hat{\Pi} (\hat{\Pi}' \hat{\Pi}^{-1}) \hat{\Pi}' ] \) is not dependent on \( \hat{\Pi} \).

In other words, \( \hat{R} \) does not vary as \( \hat{\Pi} \) varies; it is constant for all choices of \( \hat{\Pi} \), including the choice \( \hat{\Pi}=\Pi \). We saw before that if \( \hat{\Pi}=\Pi \), then \( \hat{R}\Pi = 0 \). Since \( \hat{R} \) is not dependent on the choice of \( \hat{\Pi} \), we can conclude that \( \hat{R}\Pi = 0 \) for any \( \hat{\Pi} \) used in the fitted model.

Now let us return to the bias of \( \hat{\beta}_{1s} \) and \( \hat{\gamma}_{1s} \). Notice that the
formula for \( \gamma_{ls} \), expression (3.6), is a function of three components, \( V, \hat{R}, \) and \( Y \). We have seen that the matrix \( \hat{R} \) is invariant to choices of \( \hat{\Pi} \). \( Y \) is the vector of observed responses which will obviously not change with different choices of \( \hat{\Pi} \).
Consequently, for a given matrix \( V \), the vector \( \hat{\gamma}_{ls} \) will be invariant to choices of \( \hat{\Pi} \). This result tells us that the estimated regression coefficients for the covariates have the same values for different \( \hat{\Pi} \) matrices used in the fitted model. An important implication of this result is that the vector of estimates, \( \hat{\gamma}_{ls} \), will have the same value whether or not \( \hat{\Pi}=\Pi \). If \( \hat{\Pi}=\Pi \), as we've seen before, \( \hat{\gamma}_{ls} \) is unbiased. Therefore, since the same estimates are produced for all \( \hat{\Pi} \) matrices, \( \hat{\gamma}_{ls} \) is always unbiased. This result is perhaps more clearly seen though expression (3.8), the conditional expectation of \( \hat{\gamma}_{ls} \).
Since \( \hat{R}\Pi=0 \), regardless of the value of \( \hat{\Pi} \), expression (3.8) becomes \( E(\hat{\gamma}_{ls}|\hat{\Pi})=\gamma^* \). This shows us that \( \hat{\gamma}_{ls} \) is unbiased for any \( \hat{\Pi} \) used in the fitted model. In particular, \( \hat{\gamma}_{ls} \) is an unbiased estimator of \( \gamma^* \) if \( \hat{\Pi}=\Pi^0 \), that is, if misclassification is ignored.
Looking at the expected value of \( \hat{\beta}_{ls} \), we can use the fact that \( \hat{R}\Pi=0 \) for all \( \hat{\Pi} \) to rewrite expression (3.7) as
\[
E(\hat{\beta}_{ls}|\hat{\Pi})=(\hat{\Pi}'\hat{\Pi})^{-1}\hat{\Pi}'\Pi\beta^* \tag{3.11}
\]
Substituting expression (3.9) into this expression we get
\[
E(\hat{\beta}_{1s} | \hat{\Pi}) = \frac{1}{S} \left[ (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1', (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1', \ldots, (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1' \right] \begin{bmatrix} \Pi_1 \\ \Pi_1 \\ \vdots \\ \Pi_1 \end{bmatrix} \beta^*
\]

so that \(E(\hat{\beta}_{1s} | \hat{\Pi}) = (\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1' \Pi_1 \beta^*\).

Since \(\hat{\Pi}_1\) is invertible, we can write

\[(\hat{\Pi}_1' \hat{\Pi}_1)^{-1} \hat{\Pi}_1' = \hat{\Pi}_1^{-1} .\]

This leads to the general expression

\[E(\hat{\beta}_{1s} | \hat{\Pi}) = \hat{\Pi}_1^{-1} \Pi_1 \beta^* \quad (3.12)\]

Consider a situation in which there are just two exposure categories (i.e., \(k=1\)). Using expression (3.12) we can determine the bias of \(\hat{\beta}_{1s}\). The matrix \(\hat{\Pi}_1\) is

\[
\hat{\Pi}_1 = \begin{bmatrix}
1 & \hat{\pi}_{01} \\
1 & \hat{\pi}_{11}
\end{bmatrix}
\]

so that

\[
\hat{\Pi}_1^{-1} = (\hat{\pi}_{11} - \hat{\pi}_{01})^{-1} \begin{bmatrix}
\hat{\pi}_{11} & -\hat{\pi}_{01} \\
-1 & 1
\end{bmatrix} .
\]

Also, \(\Pi_1 = \begin{bmatrix} 1 & \pi_{01} \\ 1 & \pi_{11} \end{bmatrix}\) and

\[
\beta^* = \begin{bmatrix} \beta_\alpha^* \\ \beta_1^* \end{bmatrix} .
\]

Then, \(\hat{\Pi}_1^{-1} \Pi_1 \beta^* = \begin{bmatrix} \beta_\alpha^* + \beta_1^* \frac{\hat{\pi}_{11} \hat{\pi}_{01} - \hat{\pi}_{01} \hat{\pi}_{11}}{\hat{\pi}_{11} - \hat{\pi}_{01}} \\ \beta_1^* \frac{\hat{\pi}_{11} - \hat{\pi}_{01}}{\hat{\pi}_{11} - \hat{\pi}_{01}} \end{bmatrix} .
\]
Defining $\hat{\beta}_{1s}=(\hat{\beta}_1, \hat{\beta}_1)'$, the conditional expected value of $\hat{\beta}_1$ is

$$E(\hat{\beta}_1|\hat{\Pi}) = \beta_1 \star \left( \frac{\pi_{11}-\pi_{01}}{\pi_{11}-\hat{\pi}_{01}} \right)$$

This shows us that the degree of bias involved is determined by the ratio of the true difference $(\pi_{11}-\pi_{01})$ to its estimate. If the misclassification error were ignored, the resulting matrix $\hat{\Pi}_1$ would be:

$$\hat{\Pi}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

so that $\hat{\pi}_{11}=1$ and $\hat{\pi}_{01}=0$. In this case, the expected value of $\hat{\beta}_1$ is

$$E(\hat{\beta}_1|\hat{\Pi}=\hat{\Pi}^0) = \beta_1 \star (\pi_{11}-\pi_{01}) = \beta_1 \star (\pi_{11}+\pi_{00}-1)$$

This result is similar to the result from Section 1.6.1 describing the relationship between the risk difference in the presence of misclassification and the true risk difference.

Again, by assuming a degree of misclassification such that $\pi_{11}$ and $\pi_{00}$ are both greater than $1/2$, we find bias towards the null of $\hat{\beta}_1$ when misclassification is ignored.

3.2.3 Bias of Predicted Values

As we can see below, $\hat{\Pi}\hat{\beta}_{1s}$ is invariant to changes in $\hat{\Pi}$. Using expression (3.5), we see that

$$\hat{\Pi}\hat{\beta}_{1s} = \hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'(Y-\hat{V}_{1s}')$$

Equation (3.10) tells us that $\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'$ is invariant to changes in $\hat{\Pi}$. This implies that $\hat{\Pi}\hat{\beta}_{1s}$ is also invariant to changes in $\hat{\Pi}$. An
Important consequence of this result is that \( \hat{Y}_{1s} = \Pi \hat{\beta}_{1s} + \mathbf{V} \hat{\gamma}_{1s} \), the vector of predicted responses, is invariant to changes in \( \hat{\Pi} \) for a given set of covariates. As with \( \hat{\gamma}_{1s} \), the fact that \( \hat{Y}_{1s} \) will be the same whether or not \( \Pi = \hat{\Pi} \) implies that \( \hat{Y}_{1s} \) is an unbiased estimator of the response vector when conditioning on \( \hat{\Pi} \), regardless of which \( \hat{\Pi} \) matrix is used in the fitted model.

It was shown earlier that if \( \hat{\Pi} = \Pi \), then \( \hat{\beta}_{1s} \) and \( \hat{\gamma}_{1s} \) are unbiased estimators of \( \beta^* \) and \( \gamma^* \), respectively. This unbiasedness was conditional on a fixed value of \( \hat{\Pi} \). Therefore, the unbiasedness of \( \hat{\Pi} \hat{\beta}_{1s} \) is also conditional. \( \hat{\Pi} \hat{\beta}_{1s} \), however, is also unconditionally unbiased. When determining the bias of \( \hat{Y}_{1s} \), if we treat \( \hat{\Pi} \) as a random variable, rather than a as fixed constant, we still find \( \hat{Y}_{1s} \) to be unbiased for \( E(\hat{Y}) \) for all values of \( \hat{\Pi} \). Using expression (3.12) we find

\[
E(\hat{\Pi} \hat{\beta}_{1s} | \hat{\Pi}) = \Pi \left[ E(\hat{\beta}_{1s} | \hat{\Pi}) \right] = \Pi \hat{\Pi}_{1}^{-1} \Pi_1 \beta^*
\]

\[
= \begin{bmatrix} \Pi_1 & \cdots & \Pi_k \end{bmatrix} \begin{bmatrix} \Pi_1^{-1} \Pi_1 \beta^* \\ \Pi_{k+1} \end{bmatrix} = \begin{bmatrix} I_{k+1} \\ I_{k+1} \\ \vdots \\ I_{k+1} \end{bmatrix} \Pi_1 \beta^* = \Pi \beta^*
\]

This shows us that the unconditional expected value of \( \hat{\Pi} \hat{\beta}_{1s} \) is \( \Pi \beta^* \). Further, by recalling that \( \hat{\gamma}_{1s} \) is unbiased for any \( \hat{\Pi} \), we can conclude that \( \hat{\gamma}_{1s} \) is unconditionally unbiased for \( \gamma^* \).

Consequently, the vector \( \hat{Y}_{1s} \) is an unconditionally unbiased estimator of the true vector of responses.
3.2.4 Variance-Covariance Matrices for $\hat{\beta}_{ls}$ and $\hat{\gamma}_{ls}$

Now let us examine the conditional variance of the various statistics of interest, beginning with the variance-covariance matrix of $(\hat{\beta}_{ls}, \hat{\gamma}_{ls})$. The variance of $\hat{\gamma}_{ls}$ is derived using expression (3.6) and the fact that $\hat{R}$ is symmetric and idempotent:

$$Var(\hat{\gamma}_{ls}) = (V'\hat{R}V)^{-1} V'\hat{R} \text{Var}(\gamma) \hat{R}V(V'\hat{R}V)^{-1}$$

$$= \sigma^2 (V'\hat{R}V)^{-1} (V'\hat{R}V) (V'\hat{R}V)^{-1}$$

$$= \sigma^2 (V'\hat{R}V)^{-1}$$  \hfill (3.13)

This uses the least-square assumption of a constant variance, $\sigma^2$, and mutual independence of the $\gamma_{ls}$'s, $i=0,1,...,k$, $s=1,2,...,S$.

The derivation of the covariance of $\hat{\beta}_{ls}$ and $\hat{\gamma}_{ls}$ uses expressions (3.5) for $\hat{\beta}_{ls}$ and (3.6) for $\hat{\gamma}_{ls}$.

$$\text{Cov}(\hat{\beta}_{ls}, \hat{\gamma}_{ls}) = \text{Cov}\left\{ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y, \hat{\gamma}_{ls} \right\} - \text{Cov}\left\{ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V\gamma_{ls}, \hat{\gamma}_{ls} \right\}$$

$$= \text{Cov}\left\{ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y, (V'\hat{R}V)^{-1} V'\hat{R}Y \right\} - (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V \text{Var}(\hat{\gamma}_{ls})$$

$$= \sigma^2 (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'R(V'\hat{R}V)^{-1} \hat{\Pi}'V(V'\hat{R}V)^{-1}$$

$$= - \sigma^2 (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V(V'\hat{R}V)^{-1}$$  \hfill (3.14)

since $\hat{\Pi}'\hat{R}=0$ for all $\hat{\Pi}$.

Finally,

$$\text{Var}(\hat{\beta}_{ls}) = \text{Var}\left[ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'(Y-V\hat{\gamma}_{ls}) \right]$$

$$= \text{Var}\left[ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y \right] + \text{Var}\left[ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V\hat{\gamma}_{ls} \right]$$

$$- 2\text{Cov}\left[ (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y, (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V\hat{\gamma}_{ls} \right]$$
\[
\begin{align*}
= \sigma^2 (\hat{\Pi}'\hat{\Pi})^{-1} + (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V\text{Var}(\hat{\gamma}_{1s}) V'\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1} \\
- 2\text{Cov}\left[(\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y, \hat{\gamma}_{1s}\right] V'\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1} \\
= \sigma^2 \left[(\hat{\Pi}'\hat{\Pi})^{-1} + (\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'V(V'RV)^{-1}V'\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1}\right] 
\end{align*}
\]

since we saw in the derivation of \(\text{Cov}(\hat{\beta}_{1s}, \hat{\gamma}_{1s})\) that
\[
\text{Cov}\left[(\hat{\Pi}'\hat{\Pi})^{-1} \hat{\Pi}'Y, \hat{\gamma}_{1s}\right] = 0.
\]

All three expressions, (3.13), (3.14), and (3.15), involve \(\sigma^2\), an unknown parameter. We can see from expression (3.13) that the true variance of \(\hat{\gamma}_{1s}\) will not vary as \(\hat{\Pi}\) varies. However, it is also important to see whether the estimated variance of \(\hat{\gamma}_{1s}, \text{Var}(\hat{\gamma}_{1s})\), will vary as \(\hat{\Pi}\) varies. To examine this, we must use the least squares estimate of \(\sigma^2\), which is the mean square error (MSE). MSE is equal to the error sums of squares (SSE) divided by the number of degrees of freedom corresponding to SSE, where
\[
\text{SSE} = (Y-\hat{\Pi}\hat{\beta}_{1s}-V\hat{\gamma}_{1s})'(Y-\hat{\Pi}\hat{\beta}_{1s}-V\hat{\gamma}_{1s}).
\]
We can see that \(\text{SSE}\) is invariant to changes in \(\hat{\Pi}\) since \(\hat{\Pi}\hat{\beta}_{1s}\) and \(\hat{\gamma}_{1s}\) were found to be invariant. Further, if SSE is invariant, then MSE is also invariant since the degrees of freedom will not change if the dimensions of \(\hat{\Pi}\) are not changed.

Therefore, we can conclude that both \(\text{Var}(\hat{\gamma}_{1s})\) and \(\text{Var}(\hat{\gamma}_{1s})\) are invariant to choices of \(\hat{\Pi}\). However, the true and estimated \(\text{Cov}(\hat{\beta}_{1s}, \hat{\gamma}_{1s})\) and \(\text{Var}(\hat{\beta}_{1s})\) are not invariant.
3.2.5 Variance of \( \hat{\Pi} \hat{\beta}_{ls} \)

Conditioning on \( \hat{\Pi} \), the variance of \( \hat{\Pi} \hat{\beta}_{ls} \), \( \text{Var}(\hat{\Pi} \hat{\beta}_{ls}) \), is also invariant to choices of \( \hat{\Pi} \) as the following shows:

Since

\[
\hat{\Pi}' \hat{\Pi} = S \hat{\Pi}' \hat{\Pi}_1
\]

so that

\[
(\hat{\Pi}' \hat{\Pi})^{-1} = \frac{1}{S} \hat{\Pi}_1^{-1} \hat{\Pi}_1',
\]

then expression (3.15) for the variance of \( \hat{\beta}_{ls} \) becomes

\[
\text{Var}(\hat{\beta}_{ls}) = \frac{\sigma^2}{S} \hat{\Pi}_1^{-1} \left\{ \hat{\Pi}_1' - \frac{1}{S} \hat{\Pi}_1' \left[ \hat{\Pi}_1', \hat{\Pi}_1', \ldots, \hat{\Pi}_1' \right] V(\hat{V}' \hat{R} V)^{-1} \left[ \begin{array}{c} \hat{\Pi}_1 \\ \vdots \\ \hat{\Pi}_1 \end{array} \right] \right\} \hat{\Pi}_1'^{-1}
\]

\[
= \frac{\sigma^2}{S} \hat{\Pi}_1^{-1} \left\{ I_{k+1} + \frac{1}{S} \left[ I_{k+1}, I_{k+1}, \ldots, I_{k+1} \right] V(\hat{V}' \hat{R} V)^{-1} \left[ \begin{array}{c} I_{k+1} \\ \vdots \\ I_{k+1} \end{array} \right] \right\} \hat{\Pi}_1'^{-1}
\]

\[
\Rightarrow \text{Var}(\hat{\Pi} \hat{\beta}_{ls}) = \hat{\Pi} \text{Var}(\hat{\beta}_{ls}) \hat{\Pi}'
\]

\[
= \frac{\sigma^2}{S} \left[ I_{k+1} + \frac{1}{S} \left[ I_{k+1}, I_{k+1}, \ldots, I_{k+1} \right] V(\hat{V}' \hat{R} V)^{-1} \left[ \begin{array}{c} I_{k+1} \\ \vdots \\ I_{k+1} \end{array} \right] \right]
\]

which is invariant to choices of \( \hat{\Pi} \). It also follows that the estimated variance, \( \text{Var}(\hat{\Pi} \hat{\beta}_{ls}) \), is invariant as well. Further, we
can show that the estimated and true variances of $\hat{Y}_{ls}$ are also invariant since the covariance of $\hat{N}\hat{\beta}_{ls}$ and $\gamma_{ls}$ is not a function of $\hat{N}$.

3.2.6 Invariance of $R^2$ and the F Statistic

In Section 3.2.4, we showed that SSE and MSE are invariant to choices of $\hat{N}$. The invariances of SSR and MSR, the regression sums of squares and mean squares for regression, respectively, are proven below. Through these invariances, we can show the invariance of $R^2$, used to measure goodness of fit, and of the F statistic, used to test the hypothesis that all the model coefficients are zero.

The total sums of squares, SSTO, can be partitioned into SSE and SSR (i.e., SSE+SSR=SSTO). SSR is a function of the observed and predicted values only, none of which is dependent on $\hat{N}$. SSR, and therefore MSR, are also independent of $\hat{N}$. It follows that SSTO is invariant to choices of $\hat{N}$.

$R^2$ is obtained by dividing SSR by SSTO. Since neither sums of squares is dependent on $\hat{N}$, $R^2$ does not vary with the choice of $\hat{N}$. The F statistic is the ratio of MSR to MSE. Consequently, this statistic is also invariant to choices of $\hat{N}$.

3.2.7 Summary for LS Estimation

We have shown in the preceding section that when the correct $\Pi$ matrix is used in the fitted model, $\hat{\beta}_{ls}$ and $\gamma_{ls}$ are unbiased estimators of $\beta^*$ and $\gamma^*$, respectively. When an estimate of $\Pi$, $\hat{N}$,
is used in the fitted model, $\hat{\beta}_{1s}$ is not necessarily unbiased for $\beta^*$. $\hat{\gamma}_{1s}$, on the other hand, is unbiased for $\gamma^*$ for any choice of a $\hat{\Pi}$ matrix, as $\hat{\Pi}\hat{\beta}_{1s}$ and $\hat{\gamma}_{1s}$ are for $\Pi\beta^*$ and $E(Y)$, respectively.

Also, we saw that, although $\text{Var}(\hat{\beta}_{1s})$ and $\text{Cov}(\hat{\beta}_{1s}, \hat{\gamma}_{1s})$ vary as $\hat{\Pi}$ varies, the true and estimated $\text{Var}(\hat{\gamma}_{1s})$, $\text{Var}(\hat{\Pi}\hat{\beta}_{1s})$, and $\text{Var}(\hat{Y})$ do not vary with $\hat{\Pi}$. In addition, $R^2$ and the F statistic are invariant to choices of $\hat{\Pi}$. In the next sections, we will show that similar properties hold for the WLS and ML estimators that were defined in Chapter 2.

3.3 WLS Estimation

The method of WLS can be used to estimate parameters from models (1.16) and (1.18) when $\Pi$ is known. In the case of model (1.16), an underlying binomial likelihood is assumed. In the case of model (1.18), an underlying Poisson likelihood is assumed. For the remainder of this chapter, which includes the following section dealing with ML estimators, we will limit our work to estimators which are used in Poisson regression. To work with both models would be repetitive, and the matrix algebra for the Poisson model is considerably less complex. The results reached will also hold for logistic regression.

Poisson regression methods are used to estimate the parameters in model (1.18). Model (1.18) can be written as $\ln \lambda = \Pi\beta^* + \gamma^*$. where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_k)^T$. The estimator $\hat{\lambda}_{1s}$ is unbiased for $\lambda_{1s}$, as is proven by the following:

$$E(\hat{\lambda}_{1s}) = E(X_{1s}/L_{1s}) = \left[\frac{(\lambda_{1s}L_{1s})}{L_{1s}}\right] = \lambda_{1s}.$$
Asymptotically, the estimator \( \hat{Y}_{is} = \ln(\hat{\lambda}_{is}) \) is unbiased for \( \ln(\lambda_{is}) \). Using this leads to the model \( E(Y) = \Pi \beta + V \gamma \) which, assuming model (1.18) holds, will hold in the limit. When \( \hat{\Pi} \) is used as an estimate of \( \Pi \), this model becomes \( E(Y) = \hat{\Pi} \beta + V \gamma \), where \( \beta \) and \( \gamma \) are as defined in the previous section.

3.3.1 Bias Of \( \hat{\beta}_{wls} \) And \( \hat{\gamma}_{wls} \)

Below, formulas are derived for \( \hat{\beta}_{wls} \) and \( \hat{\gamma}_{wls} \) from which we can determine their properties. This is done as follows:

Let \( Q = (Y - \hat{\Pi} \beta - V \gamma)^\prime D_x (Y - \hat{\Pi} \beta - V \gamma) \)

\[
\begin{align*}
&= Y'D_x Y - 2\beta'\hat{\Pi}'D_x Y - 2\gamma'V'D_x Y + 2\beta'\hat{\Pi}'D_x V \gamma + \beta'\hat{\Pi}'D_x \hat{\Pi} \beta \\
&
+ \gamma'V'D_x V \gamma
\end{align*}
\]

where \( D_x \) is as defined in Section 2.2.

Then,

\[
\frac{\partial Q}{\partial \beta} = -2\hat{\Pi}'D_x Y + 2\hat{\Pi}'D_x V \gamma + 2\hat{\Pi}'D_x \hat{\Pi} \beta \quad (3.16)
\]

\[
\frac{\partial Q}{\partial \gamma} = -2V'D_x Y + 2V'D_x \hat{\Pi} \beta + 2V'D_x V \gamma \quad (3.17)
\]

Setting expression (3.16) equal to 0 implies

\[
\hat{\Pi}'D_x (Y - V \hat{\gamma}_{wls}) = \hat{\Pi}'D_x \hat{\Pi} \beta_{wls}
\]

\[
\Rightarrow \hat{\beta}_{wls} = (\hat{\Pi}'D_x \hat{\Pi})^{-1} \hat{\Pi}'D_x (Y - V \hat{\gamma}_{wls}) \quad (3.18)
\]

Setting expression (3.17) equal to 0 implies, along with expression (3.18), that

\[
V'D_x \hat{\gamma}_{wls} = V'D_x Y - V'D_x \hat{\Pi} \left[(\hat{\Pi}'D_x \hat{\Pi})^{-1} \hat{\Pi}'D_x (Y - V \hat{\gamma}_{wls})\right]
\]

\[
\Rightarrow V'D_x \left[ I - \hat{\Pi}(\hat{\Pi}'D_x \hat{\Pi})^{-1} \hat{\Pi}'D_x \right] V \hat{\gamma}_{wls}
\]
\[ = V' D_x \left[ I - \hat{\Pi}(\hat{\Pi}' D_x \hat{M})^{-1} \hat{\Pi}' D_x \right] Y \]

\[ \Rightarrow V' D_x \hat{R}_w V \hat{\gamma}_{\text{wls}} = V' D_x \hat{R}_w Y \]

\[ \Rightarrow \hat{\gamma}_{\text{wls}} = \left[ V' D_x \hat{R}_w V \right]^{-1} V' D_x \hat{R}_w Y \]  \hspace{1cm} (3.19)

where \( \hat{R}_w = [ I_{S(k+1)} - \hat{\Pi}(\hat{\Pi}' D_x \hat{M})^{-1} \hat{\Pi}' D_x ] \).

As we investigated the matrix \( \hat{R} \) in Section 3.2.2, we will investigate \( \hat{R}_w \) here to gain some insight regarding the properties of \( \hat{\beta}_{\text{wls}} \) and \( \hat{\gamma}_{\text{wls}} \). Recall that \( D_x \) is a diagonal matrix with the counts \( X = (X_{01}, X_{11}, \ldots, X_{k1}, \ldots, X_{0S}, X_{1S}, \ldots, X_{kS})' \) down the diagonal. Let us write \( D_x \) as

\[
D_x = \begin{bmatrix}
D_{x1} & & \\
& D_{x2} & 0 \\
& & \ddots \\
0 & & & D_{xS}
\end{bmatrix}
\]

where \( D_{xs} \) is a diagonal matrix with the counts \( (X_{0s}, X_{1s}, \ldots, X_{ks})' \) down the diagonal. Now we can write

\[
(\hat{\Pi}' D_x \hat{M}) = \\
\left[ \hat{\Pi}_1', \hat{\Pi}_1', \ldots, \hat{\Pi}_1' \right] \left[ D_{x1} \begin{array}{c}
0 \\
D_{x2} \\
0 \\
\vdots \\
D_{xS}
\end{array} \right] \left[ \begin{array}{c}
\hat{\Pi}_1 \\
\hat{\Pi}_1 \\
\vdots \\
\hat{\Pi}_1
\end{array} \right]
\]

\[
= \left[ \hat{\Pi}_1' D_{x1} \hat{\Pi}_1 + \hat{\Pi}_1' D_{x2} \hat{\Pi}_1 + \ldots + \hat{\Pi}_1' D_{xS} \hat{\Pi}_1 \right]
\]

\[ = \hat{\Pi}_1' \left[ D_{x1} + D_{x2} + \ldots + D_{xS} \right] \hat{\Pi}_1 \]
\[
\hat{\pi}_i / \hat{G} \hat{\pi}_i
\]
where \( \hat{G} \) is a diagonal matrix with the counts
\[
\left( \sum_{s=1}^{S} X_{0s}, \sum_{s=1}^{S} X_{1s}, \ldots, \sum_{s=1}^{S} X_{ks} \right)
\]down the diagonal. Therefore \( \hat{G}^{-1} \) exists
and we have
\[
(\hat{\pi}'D_{\hat{x}}\hat{\pi})^{-1} = \hat{\pi}_i^{-1}\hat{G}^{-1}\hat{\pi}_i^{-1}
\]
Now, \( \hat{R}_w \) can be written as
\[
\hat{R}_w = I_{S(k+1)} - \left[ \begin{array}{c}
\hat{\pi}_1 \\
\vdots \\
\hat{\pi}_1
\end{array} \right] \hat{\pi}_i^{-1}\hat{G}^{-1}\hat{\pi}_i^{-1} \left[ \begin{array}{c}
\hat{\pi}_1', \\
\hat{\pi}_1', \\
\vdots \\
\hat{\pi}_1'
\end{array} \right] D_{\hat{x}}
\]
\[
= I_{S(k+1)} - \left[ \begin{array}{c}
I_{k+1} \\
I_{k+1} \\
\vdots \\
I_{k+1}
\end{array} \right] \hat{G}^{-1} \left[ \begin{array}{c}
I_{k+1}, I_{k+1}, \ldots, I_{k+1}
\end{array} \right] D_{\hat{x}}
\]
Both \( \hat{G} \) and \( D_{\hat{x}} \) are dependent only on the original counts (the \( X_{is} \)'s). We can conclude, then, that \( \hat{R}_w \) is independent of the \( \hat{\pi} \) matrix. Therefore, \( \hat{\gamma}_{wls} \) is invariant to choices of \( \hat{\pi} \).

Now, we will derive the asymptotic expectations of \( \hat{\beta}_{wls} \) and \( \hat{\gamma}_{wls} \) by first showing that they are both functions of ML estimators, and then applying asymptotic properties of ML estimators.

We begin with the likelihood of the data, which, as was mentioned in Chapter 2, is a product Poisson likelihood. This likelihood, \( L(\hat{X}) \), can be written,
\[
L(\hat{X}) = \prod_{s=1}^{S} \left( \prod_{i=0}^{k} \left[ e^{-\mu_{is}} \mu_{is}^{X_{is}} / X_{is}! \right] \right)
\]
\[
\exp \left( \sum_{s=1}^{S} \sum_{i=0}^{k} \mu_{is} \right) \prod_{s=1}^{S} \prod_{i=0}^{k} \mu_{is} \sum_{s=1}^{S} \prod_{i=0}^{k} X_{is} \prod_{s=1}^{S} \prod_{i=0}^{k} X_{is}! = \exp \left( \sum_{s=1}^{S} \prod_{i=0}^{k} X_{is} \right) \prod_{s=1}^{S} \prod_{i=0}^{k} X_{is}!.
\]

Then,
\[
\ln L(\mathbf{X}) = -\sum_{s=1}^{S} \sum_{i=0}^{k} \mu_{is} + \sum_{s=1}^{S} \sum_{i=0}^{k} X_{is} \ln \mu_{is} - \sum_{s=1}^{S} \sum_{i=0}^{k} \ln X_{is}!
\]
so that
\[
\frac{\partial \ln L(\mathbf{X})}{\partial \mu_{is}} = -1 + \frac{X_{is}}{\mu_{is}}.
\]
Setting this equal to 0 implies that the ML estimator of \( \mu_{is} \) is \( X_{is} \).

Since this holds for each \( i=0,1,\ldots,k \) and \( s=1,2,\ldots,S \), the vector \( \mathbf{X} \) is the ML estimator of the vector \( \mu=(\mu_{01},\mu_{11},\ldots,\mu_{kS})' \). In addition, \( \mathbf{D}_x \) is the ML estimator of \( \mathbf{D}_\mu \), which is a diagonal matrix with the elements of \( \mu \) down the diagonal.

In order to derive the asymptotic expectation of \( \hat{\mathbf{y}}_{wls} \), let us first return to its formula as given in expression (3.19):
\[
\hat{\mathbf{y}}_{wls} = \left[ \mathbf{V}' \mathbf{D}_x \hat{\mathbf{R}}_w \mathbf{V} \right]^{-1} \mathbf{V}' \mathbf{D}_x \hat{\mathbf{R}}_w \mathbf{Y}.
\]
\( \mathbf{Y} \) can be written as \( \ln(\mathbf{D}_L^{-1} \mathbf{X}) \) where \( \ln \) refers to a function which takes the natural log of each element in a matrix, and \( \mathbf{D}_L \) is a diagonal matrix with the amounts of population-time \( \mathbf{L}=(L_{01},L_{11},\ldots,L_{kS})' \) down the diagonal.

By the large-sample properties of ML estimators, we can say that the asymptotic expected value of \( \hat{\mathbf{y}}_{wls} \) is
\[
\mathbb{E}_a(\hat{\mathbf{y}}_{wls} | \hat{\mathbf{M}}) = \left\{ \left[ \mathbf{V}' \mathbf{D}_x \hat{\mathbf{R}}_w \mathbf{V} \right]^{-1} \mathbf{V}' \mathbf{D}_x \hat{\mathbf{R}}_w \ln(\mathbf{D}_L^{-1} \mu) \right\}
\]
Substituting \( \ln(\mathbf{D}_L^{-1} \mu) = \ln \lambda = \Pi \mathbf{b}^* + \mathbf{y} \mathbf{\gamma}^* \) into the above expression
leads to

\[ E_a(\hat{\gamma}_{\text{wls}} \mid \hat{\Pi}) = \left\{ \left[ V'D_{\mu} \hat{R}_w V \right]^{-1} V'D_{\mu} \hat{R}_w (\Pi\beta^* + V\gamma^*) \right\} \]

Notice that \( \hat{R}_w \Pi = 0 \) for any \( \hat{\Pi} \); then, the invariance of \( \hat{R}_w \) implies that \( \hat{R}_w \Pi = 0 \) for any \( \hat{\Pi} \). Using this, we find

\[ E_a(\hat{\gamma}_{\text{wls}} \mid \hat{\Pi}) = E_a(\gamma_{\text{wls}}) = \gamma^* \quad (3.20) \]

Expression (3.20) indicates that \( \hat{\gamma}_{\text{wls}} \) is, asymptotically, an unbiased estimator of \( \gamma^* \) for any \( \hat{\Pi} \) matrix.

The conditional asymptotic expectation of \( \hat{\beta}_{\text{wls}} \) can be derived from expression (3.18).

\[ E_a(\hat{\beta}_{\text{wls}} \mid \hat{\Pi}) = E_a \left\{ (\hat{\Pi}'D_{x'}\hat{\Pi})^{-1} \hat{\Pi}'D_{x'} \left( \ln(D_{L_j}^{-1}x) - V\gamma_{\text{wls}} \right) \mid \hat{\Pi} \right\} \]

\[ = (\hat{\Pi}'D_{x'}\hat{\Pi})^{-1} \hat{\Pi}'D_{x'} \left( \Pi\beta^* + V\gamma^* - V\gamma^* \right) \]

since we saw above that \( E_a(\gamma_{\text{wls}} \mid \hat{\Pi}) = E_a(\gamma_{\text{wls}}) = \gamma^* \).

As we proved the identity \( (\hat{\Pi}'D_{x'}\hat{\Pi})^{-1} = \hat{\Pi}_i^{-1} G \hat{\Pi}_i^{-1} \),

we can also show that \( (\hat{\Pi}'D_{\mu}\hat{\Pi})^{-1} = \hat{\Pi}_i^{-1} G \hat{\Pi}_i^{-1} \), where \( G \) is a diagonal matrix with the counts \( \sum_{s=1}^S \mu_0, \sum_{s=1}^S \mu_1, \ldots, \sum_{s=1}^S \mu_L \) down the diagonal. This leads us to

\[ E_a(\hat{\beta}_{\text{wls}} \mid \hat{\Pi}) = \hat{\Pi}_i^{-1} G^{-1} \hat{\Pi}_i^{-1} \left[ \hat{\Pi}_i', \hat{\Pi}_i', \ldots, \hat{\Pi}_i' \right] D_{\mu} \Pi\beta^* \]

\[ = \hat{\Pi}_i^{-1} G^{-1} \left[ I_{k+1}, I_{k+1}, \ldots, I_{k+1} \right] D_{\mu} \Pi\beta^* \]

\[ = \hat{\Pi}_i^{-1} G^{-1} \left[ D_{\mu_1}, D_{\mu_2}, \ldots, D_{\mu_S} \right] \left[ \begin{array}{c} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_i \end{array} \right] \beta^* \]
\[ = \hat{\Pi}_1^{-1} G^{-1} \left[ D_{\mu_1} \Pi_1 + D_{\mu_2} \Pi_1 + \ldots + D_{\mu_s} \Pi_1 \right] \beta^* \]

\[ = \hat{\Pi}_1^{-1} G \Pi_1 \beta^* \]

\[ = \hat{\Pi}_1^{-1} \Pi_1 \beta^* \]

where \( D_{\mu_s} \) is a diagonal matrix with the elements \( (\mu_0, \mu_1, \ldots, \mu_k) \) down the diagonal.

The above is identical to expression (3.12) for the conditional expectation of \( \hat{\beta}_s \). Clearly, if \( \hat{\Pi} = \Pi \), \( \hat{\beta}_{\text{wls}} \) is a conditionally unbiased estimator of \( \beta^* \) in the limit. Also, as we saw with the LS estimator, if misclassification is ignored in a situation involving just two exposure categories, \( \hat{\beta}_1 \) will be biased towards the null.

### 3.3.2 Bias of Predicted Values

The vector of predicted values, \( \hat{Y}_{\text{wls}} = \hat{\Pi} \hat{\beta}_{\text{wls}} + \hat{V}_Y \), is invariant as it was in the case of LS estimation. Again, it hinges on the fact that \( \hat{\Pi} \hat{\beta}_{\text{wls}} \) is invariant to choices of \( \hat{\Pi} \), which is shown below. From expression (3.18), it follows that

\[ \hat{\Pi} \hat{\beta}_{\text{wls}} = \hat{\Pi} (\hat{\Pi}'D_x \hat{\Pi})^{-1} \hat{\Pi}'D_x (Y - \hat{V}_Y) \]

\[ = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_1 \hat{G}^{-1} \hat{\Pi}_1' \\ \vdots \\ \hat{\Pi}_1 \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1' \hat{\Pi}_1, \hat{\Pi}_1', \ldots, \hat{\Pi}_1' \end{bmatrix} D_x (Y - \hat{V}_Y)_{\text{wls}} \]
\[
\begin{bmatrix}
I_{k+1} & \hat{I}_{k+1} & \hat{I}_{k+1} & \ldots & \hat{I}_{k+1}
\end{bmatrix}
\mathbf{G}^{-1}
\begin{bmatrix}
I_{k+1} & I_{k+1} & \ldots & I_{k+1}
\end{bmatrix}
\mathbf{D}_x \left( Y - \hat{Y}_{\text{wls}} \right)
\]

which is not a function of \( \hat{\Pi} \). Further, this implies that the vector of predicted counts, \( \hat{\mu}_{\text{wls}} \), is also invariant.

Now, let us examine the bias of \( \hat{\Pi}\hat{\beta}_{\text{wls}} \). The conditional asymptotic expected value of \( \hat{\Pi}\hat{\beta}_{\text{wls}} \) can be written

\[
E(\hat{\Pi}\hat{\beta}_{\text{wls}} | \hat{\Pi}) = \hat{\Pi} E(\hat{\beta}_{\text{wls}} | \hat{\Pi}) = \hat{\Pi}(\hat{\Pi}_1^{-1}\Pi_i\beta^*)
\]

\[
= \begin{bmatrix}
\hat{\Pi}_1 \\
\hat{\Pi}_1 \\
\vdots \\
\hat{\Pi}_1
\end{bmatrix}
\begin{bmatrix}
I_{k+1} & I_{k+1} & \ldots & I_{k+1}
\end{bmatrix}
\Pi_i\beta^* = \Pi\beta^*
\]

This implies that, asymptotically, \( \hat{\Pi}\hat{\beta}_{\text{wls}} \) is an unconditionally unbiased estimator of \( \Pi\beta^* \), and therefore \( \hat{Y}_{\text{wls}} \) is an unconditionally unbiased estimator of \( E(Y) \).

3.3.3 Goodness-of-fit Statistic for WLS Estimation

Goodness of fit for the WLS estimates is measured using the Wald goodness-of-fit chi-square statistic,

\[
Q_w = \left[ \ln \hat{\lambda} - \hat{Y}_{\text{wls}} \right]' \mathbf{D}_x \left[ \ln \hat{\lambda} - \hat{Y}_{\text{wls}} \right].
\]

The vector \( \hat{\lambda} \), defined in Section 2.2, and the matrix \( \mathbf{D}_x \) are dependent only on the original data. \( \hat{Y}_{\text{wls}} \), we have seen, is invariant to choices of \( \hat{\Pi} \). Therefore, \( Q_w \) will be independent of \( \hat{\Pi} \).
The same goodness-of-fit test will result regardless of the $\hat{\Pi}$ matrix used in the fitted model.

3.3.4 Variance-Covariance Matrices for $\hat{\beta}_{wls}$ and $\hat{\gamma}_{wls}$

The estimated asymptotic variance-covariance matrix of $\hat{\beta}_{wls}$ and $\hat{\gamma}_{wls}$ is

$$\text{Var}_a(\hat{\beta}_{wls}, \hat{\gamma}_{wls}) = \begin{bmatrix} \hat{\Pi}' & D_x \hat{\Pi} \end{bmatrix} \begin{bmatrix} \hat{\Pi} \hat{\gamma} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}' \hat{\Pi} \end{bmatrix}.$$

By multiplying out

$$\begin{bmatrix} \hat{\Pi}' \\ \hat{\Pi} \end{bmatrix} D_x \begin{bmatrix} \hat{\Pi} \hat{\gamma} \end{bmatrix}$$

and using techniques to invert partitioned matrices, we can arrive at expressions for the estimated asymptotic variances and covariance: $\text{Var}_a(\hat{\beta}_{wls})$, $\text{Cov}_a(\hat{\beta}_{wls}, \hat{\gamma}_{wls})$, and $\text{Var}_a(\hat{\gamma}_{wls})$. The estimated asymptotic variance of $\hat{\beta}_{wls}$ can be written as

$$\text{Var}_a(\hat{\beta}_{wls}) = \hat{\Pi}_{\Pi}^{-1} H \hat{\Pi}_{\Pi}^{-1}$$

where $H$ is a matrix whose elements are solely functions of the original counts, $X=(X_{01}, X_{11}, \ldots, X_{kS})'$. Therefore, we can see that $\text{Var}_a(\hat{\beta}_{wls})$ does vary with the choice of $\hat{\Pi}$.

Likewise, $\text{Cov}_a(\hat{\beta}_{wls}, \hat{\gamma}_{wls})$ is dependent on $\hat{\Pi}$. $\text{Var}_a(\hat{\gamma}_{wls})$, however, is not a function of $\hat{\Pi}$. It is a function only of $\hat{\Pi}$ and the original counts. Therefore, $\text{Var}_a(\hat{\gamma}_{wls})$ will not vary as $\hat{\Pi}$ varies.

3.3.5 Summary for WLS Estimation

In this section, we found the following asymptotic properties of the estimators. If $\hat{\Pi}=\Pi$, then $\hat{\beta}_{wls}$ is a conditionally unbiased
estimator of $\beta^*$. Further, $\gamma_{wls}$, $\Pi_\beta_{wls}$, and $\gamma_{wls}$ are unconditionally unbiased for $\gamma^*$, $\Pi_\beta^*$, and $E(Y)$, respectively, regardless of the value of $\hat{\Pi}$. We also determined that, although $\text{Var}_{a}(\hat{\beta}_{wls})$ and $\text{Cov}_{a}(\hat{\beta}_{wls}, \gamma_{wls})$ vary with the choice of $\hat{\Pi}$, $\text{Var}_{a}(\gamma_{wls})$ is independent of $\hat{\Pi}$. In addition, the Wald goodness-of-fit statistic is invariant to choices of $\hat{\Pi}$. Each of these results also holds in the case of logistic regression.

3.4 ML Estimation

As with the WLS estimators, we will be dealing with ML estimators in the context of Poisson regression. The results found here also hold for logistic regression. The two models of interest are identical to the ones presented in the previous section, namely: the true model, $E(\ln \lambda) = E(Y) = \Pi_\beta^* + V\gamma^*$, and the model to be fitted, $E(\ln \lambda) = E(Y) = \hat{\Pi}_\beta + V\gamma$. Again, these models make use of the property that the asymptotic expected value of $Y$ is $\ln \lambda$.

3.4.1 Bias of $\hat{\beta}_{m1}$ and $\hat{\gamma}_{m1}$

Recall from Chapter 2 that the maximum likelihood estimators, $\hat{\beta}_{m1}$ and $\hat{\gamma}_{m1}$, are arrived at through an iteration process for which the beginning value is the WLS estimate. Therefore, the first stage values of the procedure are

$$
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\gamma}_1
\end{pmatrix} = 
\begin{pmatrix}
\hat{\beta}_{wls} \\
\hat{\gamma}_{wls}
\end{pmatrix} + 
\begin{pmatrix}
\Pi' \\
V'
\end{pmatrix}
D_{\mu_{wls}} [\Pi, V]^{-1}
[\Pi'] (X - \mu_{wls})
$$
where \( \hat{\mu}_{\text{wls}} \) is the vector of predicted counts from the WLS estimation and \( \hat{D} \) is a diagonal matrix with these predicted counts down the main diagonal.

Multiplying out this matrix expression, we find

\[
\begin{bmatrix}
\hat{\Pi}' \\
\hat{V}'
\end{bmatrix}
\hat{D}_{\mu_{\text{wls}}} \begin{bmatrix}
\hat{\Pi} \\
\hat{V}
\end{bmatrix} =
\begin{bmatrix}
\hat{\Pi}'D_{\mu_{\text{wls}}}\hat{\Pi} & \hat{\Pi}'D_{\mu_{\text{wls}}}\hat{V} \\
V'D_{\mu_{\text{wls}}}\hat{\Pi} & V'D_{\mu_{\text{wls}}}\hat{V}
\end{bmatrix}.
\]

This matrix can be inverted using techniques to invert a partitioned matrix. Define the resulting matrix as the partitioned matrix

\[
\left\{ \begin{bmatrix}
\hat{\Pi}' \\
\hat{V}'
\end{bmatrix}
\hat{D}_{\mu_{\text{wls}}} \begin{bmatrix}
\hat{\Pi} \\
\hat{V}
\end{bmatrix} \right\}^{-1} =
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}.
\]

Using the inversion techniques, matrix expressions are arrived at for each of the matrices \( A_{11}, A_{12}, A_{21}, A_{22} \). We find that \( A_{11} \) can be written as \( A_{11} = \hat{\Pi}_i^{-1} H_i \hat{\Pi}_i^{-1} \), \( A_{12} \) as \( A_{12} = -\hat{\Pi}_i H_i H_4 V A_{22} \), \( A_{21} \) as \( A_{21} = -A_{22} V' H_3 \hat{\Pi}_i^{-1} \), and \( A_{22} \) as \( A_{22} = (V'H_4 V)^{-1} \) where \( H_1, H_2, H_3, \) and \( H_4 \) are matrices whose elements are solely functions of \( \hat{\mu}_{\text{wls}} \).

Now the formulas for \( \hat{\beta}_1 \) and \( \hat{\gamma}_1 \) become

\[
\hat{\beta}_1 = (A_{11}\hat{\Pi} + A_{12}V') (X - \hat{\mu}_{\text{wls}}) + \hat{\beta}_{\text{wls}}
\]

and

\[
\hat{\gamma}_1 = (A_{21}\hat{\Pi} + A_{22}V') (X - \hat{\mu}_{\text{wls}}) + \hat{\gamma}_{\text{wls}}
\]

Let us investigate the expression for \( \hat{\gamma}_1 \) first. Notice that, since we showed in the previous section that \( \hat{\mu}_{\text{wls}} \) is invariant
to choices of \( \hat{\Pi} \), the matrix \( A_{22} \) is invariant to choices of \( \hat{\Pi} \). Also, when expanding

\[
A_{21} \hat{\Pi}' = -A_{22} V'H_3 \hat{\Pi}_1^{-1} \left[ \hat{\Pi}_1', \hat{\Pi}_1', \ldots, \hat{\Pi}_1' \right]
\]

\[
= -A_{22} V'H_3 \left[ I_{k+1}, I_{k+1}, \ldots, I_{k+1} \right]
\]

we see that \( A_{21} \hat{\Pi}' \) is invariant to choices of \( \hat{\Pi} \) as well. This, along with the fact that \( \hat{\gamma}_{wls} \) is invariant, leads us to conclude that \( \hat{\gamma}_1 \) is also invariant.

The formula for \( \hat{\beta}_1 \) shows us that \( \hat{\beta}_1 \) is not invariant to changes in \( \Pi \). However, if we examine the estimator \( \hat{\Pi}\hat{\beta}_1 \), we find

\[
\hat{\Pi}\hat{\beta}_1 = \hat{\Pi}^{-1} \left[ H_1 \hat{\Pi}_1^{-1} \hat{\Pi}' - H_2 V^{-1} V' \right] (X - \hat{\mu}_{wls}) + \hat{\Pi}\hat{\beta}_{wls}
\]

\[
= \begin{bmatrix} I_{k+1} & I_{k+1} & \ldots & I_{k+1} \end{bmatrix} (H_1[I_{k+1}, I_{k+1}, \ldots, I_{k+1}] - H_2) (X - \hat{\mu}_{wls}) + \hat{\Pi}\hat{\beta}_{wls}
\]

which is invariant to choices of \( \hat{\Pi} \). An important consequence of this result is that the vector of predicted counts, \( \hat{\mu}_1 = \hat{\Pi}\hat{\beta}_1 + V\hat{\gamma}_1 \), is also invariant.

Therefore, we have seen that for the first stage of the iteration procedure, \( \hat{\gamma}_1 \) and \( \hat{\mu}_1 \) are invariant to choices of \( \hat{\Pi} \). Now, let us investigate the properties of these two estimators for successive stages. The formulas for stage two are

\[
\hat{\beta}_2 = (B_{11}\hat{\Pi}' + B_{12}V') (X - \hat{\mu}_1) + \hat{\beta}_1
\]

and
\[ \hat{\gamma}_2 = (B_{11} \Pi' + B_{22} V') (x - \hat{\mu}_1) + \hat{\gamma}_1 \]

where \( B_{11} = \Pi_1^{-1} M_1 \Pi_1^{-1} \), \( B_{12} = -\Pi_1^{-1} M_2 V' \), \( B_{21} = -B_{22} V'M_3 \Pi' \), \( B_{22} = (V'M_4 V)^{-1} \) and \( M_1, M_2, M_3, \) and \( M_4 \) are matrices whose elements are functions solely of \( \hat{\mu}_1 \). Clearly, \( \hat{\gamma}_2, \Pi \beta_2 \), and therefore \( \hat{\mu}_2 \), will be invariant to choices of \( \hat{\Pi} \). Also, it is obvious that this invariance will hold at each stage of the iteration procedure, including the final stage which produces \( \hat{\beta}_{m1} \) and \( \hat{\gamma}_{m1} \).

The invariance of \( \hat{\gamma}_{m1} \) indicates that it is asymptotically unbiased for \( \gamma^* \). This is shown be the following argument. If \( \hat{\Pi} = \Pi \) (i.e., the true model is fitted), then, by the large-sample properties of ML estimators, \( \hat{\gamma}_{m1} \) is asymptotically unbiased for \( \gamma^* \). Since \( \hat{\gamma}_{m1} \) is invariant, the vector \( \hat{\gamma}_{m1} \) will be identical for each choice of \( \hat{\Pi} \), including the choice \( \hat{\Pi} = \Pi \). Therefore, \( \hat{\gamma}_{m1} \) is always asymptotically unbiased for \( \gamma^* \). The same argument can be used to show that \( \hat{\mu}_{m1} \) is asymptotically unbiased for \( \mu \).

Since \( \hat{\beta}_{m1} \) is not invariant, it is only asymptotically unbiased for \( \beta^* \) when \( \hat{\Pi} = \Pi \).

3.4.2 Variance-Covariance Matrices for \( \hat{\beta}_{m1} \) and \( \hat{\gamma}_{m1} \)

The asymptotic variance-covariance matrix of the ML estimators is estimated at the first stage of the iteration by

\[
\text{Var}_a(\hat{\beta}_1, \hat{\gamma}_1) = \left[ \begin{bmatrix} \Pi' \\ V' \end{bmatrix} D_{\hat{\mu}_1} \begin{bmatrix} \hat{\Pi}, V \end{bmatrix} \right]^{-1}
\]

As we have just seen, \( \hat{\mu}_1 \) will be invariant to \( \hat{\Pi} \). Therefore, by
examining the structure of the above matrix, whose components are presented in the previous subsection, we can see that $\text{Var}_a(\hat{\gamma}_i)$ alone will be invariant to $\hat{\Pi}$. Clearly, this is true at all other stages of the procedure as well.

3.4.3 Goodness-of-fit Statistics for ML Estimation

In Section 2.3, two chi-square statistics were said to be used for testing goodness of fit with ML estimation. The formulas for these statistics are given below:

$$Q_p = \sum_{s=1}^{S} \sum_{i=0}^{k} \frac{(X_{is} - \hat{\mu}_{is})^2}{\hat{\mu}_{is}}$$

$$Q_{log} = \sum_{s=1}^{S} \sum_{i=0}^{k} 2X_{is} \left[ \ln(X_{is}/\hat{\mu}_{is}) \right]$$

Notice that both these statistics are functions solely of the observed counts and the predicted counts. As we have seen, the predicted counts for ML estimation are not dependent on $\hat{\Pi}$. Therefore, we can conclude that both $Q_p$ and $Q_{log}$ are invariant to choices of $\hat{\Pi}$.

3.4.4 Summary for ML Estimation

As with the LS and WLS cases, we showed that $\hat{\gamma}_{ml}$ and the predicted counts are invariant to choices of $\hat{\Pi}$. From this, it follows that $\hat{\gamma}_{ml}$ and $\hat{\mu}_{ml}$ are asymptotically unbiased for $\gamma^*$ and $\mu$, respectively. $\hat{\beta}_{ml}$, on the other hand, is not invariant and will only be asymptotically unbiased for $\beta^*$ when $\hat{\Pi} = \Pi$. We also saw that
the estimated asymptotic variance of $\hat{\gamma}_{ml}$ is invariant to $\hat{\Pi}$, while the estimated asymptotic variance of $\hat{\beta}_{ml}$ and covariance of $\hat{\beta}_{ml}$ and $\hat{\gamma}_{ml}$ are not. Further, $Q_p$ and $Q_{log}$, the goodness-of-fit statistics, were found to be invariant to choices of $\hat{\Pi}$. 
CHAPTER 4
EXTENSIONS AND IMPLICATIONS OF RESULTS

4.1 Introduction

In the previous chapter, many results were found concerning the properties of several types of estimators in a situation involving misclassification of exposure with equal misclassification probabilities over all strata. In this chapter, we will examine the implications of these results on other types of situations involving misclassification error. For instance, we will look at how the results from Chapter 3 can be applied to situations involving misclassification of covariates. We will also examine the effect of misclassification probabilities which differ over strata.

In addition, a relationship is derived between the estimate of $\beta$ obtained when misclassification is ignored and the estimate obtained when the correct $\Pi$ matrix is used in the fitted model. From this relationship, several results are shown concerning the properties of the estimate of $\beta$ obtained when misclassification is ignored.
4.2 Misclassification of Covariates

In this section, several results are derived for properties of estimators when there is misclassification of covariates, rather than exposure. The results found here are similar to those found in Chapter 3. In fact, the derivations used in Chapter 3 are frequently applied to this altered scenario.

To begin, consider again the example presented in Section 2.5.2, excluding the race factor. There were six strata involved, defined by three age and two sex categories. The matrix representation of the model

$$E(Y) = \Pi^0 \beta^* + \Phi \gamma^*$$

(4.1)

is

$$\Pi^0 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \beta^* = \begin{pmatrix} \beta_{1*} \\ \beta_{2*} \end{pmatrix}, \quad \Phi = \begin{pmatrix} \psi_{12} & \psi_{13} & \xi_{12} \\ \psi_{12} & \psi_{13} & \xi_{12} \\ \psi_{22} & \psi_{23} & \xi_{12} \\ \psi_{22} & \psi_{23} & \xi_{12} \\ \psi_{32} & \psi_{33} & \xi_{12} \\ \psi_{32} & \psi_{33} & \xi_{12} \end{pmatrix}$$

$$\gamma^* = \begin{pmatrix} \gamma_{1*} \\ \gamma_{2*} \\ \gamma_{3*} \end{pmatrix}$$

where $\beta_{1*}$, the intercept term, is the reference value for the group with no exposure in the first age and sex categories. $\beta_{1*}$ is the exposure effect. $\gamma_{1*}$ and $\gamma_{2*}$ are the effects for the second and third age groups, respectively. $\gamma_{3*}$ is the effect for the second sex.
category.

Now, redefine this model by including the intercept term in the vector of covariate coefficients. This new model, whose matrix components are presented below, is equivalent to model (4.1). The new model is written

$$E(\gamma) = \tilde{\Pi}^0 \tilde{\beta}^* + \Phi \tilde{\gamma}^*$$

(4.2)

with

$$\tilde{\beta}^* = (\beta_{1}^*) \quad , \quad \Phi \tilde{V} = [1, \Phi V] \quad , \quad \tilde{\gamma}^* = \begin{pmatrix} \beta_{\alpha}^* \\ \gamma_{1}^* \\ \gamma_{2}^* \\ \gamma_{3}^* \end{pmatrix}$$

where 1 is a column of 1's, and

$$\tilde{\Pi}^0 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\tilde{V}$$ is defined as $$\tilde{V} = [1, V]$$. From the definition of $$\Phi$$ given in Section 2.5.1 and the fact that \(\sum \phi_{sr} = 1\), it follows that $$\Phi \tilde{V} = [1, \Phi V]$$.

The motivation behind this new representation of model (4.1) is the following. Model (4.2) is equivalent to model (4.1); however, the intercept parameter in (4.2) is housed with the covariate coefficients.
This is done in order to create a model which is an exact analogy of the model developed for misclassification of exposure,

\[ E(\gamma) = \Pi \beta^* + V \gamma^*. \]

In this model, \( \Pi \) is a matrix whose first column is \( \mathbf{1} \), and whose remaining columns contain elements which are misclassification probabilities. The same can be said of \( \tilde{\Phi} \tilde{V} \). \( \beta^* \) is the vector of parameters, including the reference parameter, which is associated with the columns of \( \Pi \). \( \gamma^* \) is the vector of parameters, including the reference parameter, which is associated with the columns of \( \tilde{\Phi} \tilde{V} \). \( V \) is a matrix representing perfectly classified variables. The same can be said of \( \tilde{\Pi}^0 \). \( \gamma^* \) is the vector of parameters associated with the columns of \( V \). \( \tilde{\beta}^* \) is the vector of parameters associated with the columns of \( \tilde{\Pi}^0 \). The role played by \( \Pi \) in Chapter 3 is now played by \( \tilde{\Phi} \tilde{V} \); the roles played by \( \beta^* \), \( V \), and \( \gamma^* \) are now played by \( \tilde{\gamma}^* \), \( \tilde{\Pi}^0 \), and \( \tilde{\beta}^* \), respectively.

Earlier, we examined properties of estimators when an estimate of \( \Pi \) is used in the fitted model. Here, we will examine the properties of estimators when an estimate of \( \tilde{\Phi} \) is used in the fitted model. The new model of interest is the model to be fitted:

\[ E(\gamma) = \tilde{\Pi}^0 \tilde{\beta} + \tilde{\Phi} \tilde{V} \tilde{\gamma}. \]

In Chapter 3, the fact that \( \hat{\Pi} \) is a series of vertically concatenated matrices which are identical and invertible led to the invariance of \( \hat{R} \) and \( \hat{R}_w \). This, in turn, led to the invariance of \( \hat{\gamma} \) for LS and WLS estimation. If \( \hat{\Phi} \hat{V} \) were a series of vertically concatenated identical invertible matrices, \( \hat{\beta} \) would be invariant for choices of \( \hat{\Phi} \) for LS and
WLS estimation. However, as the example above illustrates, \( \hat{\Phi}V \) will not necessarily have this property. Therefore, the derivations in Chapter 3 cannot be used to prove the invariance of the matrix expressions \( \hat{\Phi}V' \left( \hat{\Phi}V \right)'^{-1} \left( \hat{\Phi}V \right)' \) and \( \hat{\Phi}V' \left( \hat{\Phi}V \right)'D_x \left( \hat{\Phi}V \right)'^{-1} \left( \hat{\Phi}V \right)'D_x \), which are the analogies to \( R \) and \( R_{\hat{v}} \) under misclassification of covariates.

Surprisingly, though, these expressions are invariant to choices of \( \hat{\Phi} \). This is due to the fact that for each \( \hat{\Phi} \), there exists a matrix \( T \) such that \( \hat{\Phi}V = \tilde{V}T \) and \( T \) is invertible. Consider the matrix \( \Phi V \) presented in the above example. An estimate of the corresponding \( \hat{\Phi}V \) matrix has \( \tilde{V} \) and \( T \) matrices as follows:

\[
\tilde{V} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 \\
\end{bmatrix}
\]

and \( T = \begin{bmatrix}
1 & \hat{\psi}_{12} & \hat{\psi}_{13} & \xi_{12} \\
0 & \hat{\psi}_{22} - \hat{\psi}_{12} & \hat{\psi}_{23} - \hat{\psi}_{13} & 0 \\
0 & \hat{\psi}_{32} - \hat{\psi}_{12} & \hat{\psi}_{33} - \hat{\psi}_{13} & 0 \\
0 & 0 & 0 & \xi_{22} - \xi_{12} \\
\end{bmatrix} \)

Using this equality, the expression \( \hat{\Phi}V' \left( \hat{\Phi}V \right)'^{-1} \left( \hat{\Phi}V \right)' \) becomes

\[
\tilde{V}T'\tilde{V}'\tilde{V}^{-1}T'\tilde{V} = \tilde{V}[\tilde{V}'\tilde{V}]^{-1}\tilde{V}'.
\]  

(4.3)

Similarly, the expression \( \hat{\Phi}V' \left( \hat{\Phi}V \right)'D_x \left( \hat{\Phi}V \right)'^{-1} \left( \hat{\Phi}V \right)'D_x \) becomes

\[
\tilde{V}T'\tilde{V}'D_x^{-1}T'\tilde{V}'D_x = \tilde{V}[\tilde{V}'D_x^{-1}\tilde{V}'D_x].
\]  

(4.4)

Expressions (4.3) and (4.4) are analogous to expressions \( \hat{R} \) and

94
\( \hat{R}_w \) defined under misclassification of exposure. Both (4.3) and (4.4) are not dependent on \( \hat{\phi} \). Using arguments analogous to those in Chapter 3, we can conclude that \( \hat{\beta} \) is invariant to \( \hat{\phi} \) for LS and WLS estimation.

The predicted counts for WLS estimation are also invariant. This follows from the invariance of \( (\hat{\phi} \hat{\gamma}^*) \) which is proven as follows. Expression (3.18) for \( \hat{\beta}_{wls} \) for model (4.1) can be used to determine the formula for \( \hat{\gamma}_{wls} \) for model (4.2). This is accomplished simply by replacing \( \hat{\Pi} \) by \( \hat{\phi} \hat{\gamma} \), \( \hat{V} \) by \( \hat{\Pi}^0 \), and \( \hat{\gamma}_{wls} \) by \( \hat{\beta}_{wls} \). The resulting formula is

\[
\hat{\gamma}_{wls} = [(\hat{\phi} \hat{\gamma})' D_X (\hat{\phi} \hat{\gamma})]^{-1} (\hat{\phi} \hat{\gamma})' D_X (Y - \hat{\Pi}^0 \hat{\beta}_{wls}).
\]

Using the equality \( \hat{\phi} \hat{\gamma} = \hat{\gamma} \hat{T} \) where \( \hat{T} \) is invertible leads to

\[
\hat{\phi} \hat{\gamma}_{wls} = \hat{\gamma} \hat{T} [T' \hat{\gamma} D_X \hat{T}]^{-1} T' \hat{\gamma} D_X (Y - \hat{\Pi}^0 \hat{\beta}_{wls})
\]

\[
= \hat{\gamma} [\hat{\gamma} D_X \hat{T}]^{-1} \hat{\gamma} D_X (Y - \hat{\Pi}^0 \hat{\beta}_{wls}).
\]

This expression is not dependent on the choice of \( \hat{\phi} \). Therefore, \( \hat{\gamma}_{wls} = (\hat{\Pi}^0 \hat{\beta}_{wls} + \hat{\phi} \hat{\gamma}_{wls}) \), and in turn \( \hat{\mu}_{wls} \), are invariant to choices of \( \hat{\phi} \).

\( \hat{\beta}_{m1} \) is also invariant to \( \hat{\phi} \). This is shown using the formula in Section 3.4.1 for the first stage values of the iteration procedure, namely

\[
\hat{\beta}_1 = (A_{11} \hat{\Pi}^0 + A_{12} \hat{V}) (X - \hat{\mu}_{wls}) + \hat{\beta}_{wls},
\]

where \( A_{11} = \Pi_1^{-1} H_1 \hat{\Pi}_1, A_{12} = -\Pi_1 H_2 (V' H_4 V)^{-1} \) and \( H_1, H_2 \) and \( H_4 \) are matrices whose elements are solely functions of \( \hat{\mu}_{wls} \). This formula for \( \hat{\beta} \) applies to situations in which model (1.16) is being
fitted. Since model (4.2), rather than (1.16), is used in the context of covariate misclassification, in order to arrive at a formula for \( \hat{\beta} \) the matrix \( \hat{\Pi} \) in the above expression is replaced by \( \hat{\Pi}^0, \hat{\Pi}_1 \) by \( \hat{\Pi}^0_1 \), and \( \Psi \) by \( \hat{\Theta} \).

After making these substitutions and replacing \( A_{11} \) and \( A_{12} \) with their definitions, the formula for \( \hat{\beta}_1 \) becomes

\[
\hat{\beta}_1 = (\hat{\Pi}^0_1 \hat{H}_1 \hat{\Pi}^0_1 \hat{H}^0 \hat{\Pi}^0_1 \hat{H}_2 (\hat{\Psi} \hat{\Phi}^0) \hat{\Psi} \hat{\Phi}^0) (X - \hat{\mu}_{wls}) + \hat{\beta}_{wls}
\]

\[
= (\hat{\Pi}^0_1 \hat{H}_1 \hat{\Pi}^0_1 \hat{H}^0 \hat{\Pi}^0_1 \hat{H}_2) (X - \hat{\mu}_{wls}) + \hat{\beta}_{wls}
\]

which is independent of \( \hat{\Theta} \). The expression \( (\hat{\Theta} \hat{\Psi}_1 \hat{\Phi}^0) \) can also be shown to be invariant. This leads to the invariance of \( \hat{\mu}_1 \), and of later stage iteration values of \( \hat{\beta} \), including \( \hat{\beta}_{ml} \).

In summary, we have shown that in the context of misclassified covariates, when the estimated degree of misclassification varies, the estimates of \( \hat{\beta} \) remain the same for LS, WLS, and ML estimation. In particular, if the chosen \( \hat{\Theta} \) matrix is, in fact, the true \( \Theta \) matrix, the value of \( \hat{\beta} \) will be the same as if another \( \hat{\Theta} \) matrix were chosen. This implies that \( \hat{\beta} \) is an unbiased, or asymptotically unbiased in the case of WLS and ML estimation, estimator of \( \tilde{\beta} \). In addition, the predicted counts, and therefore the goodness-of-fit statistics will be invariant to choices of \( \hat{\Theta} \) for all three types of estimation.

Below we will argue that the invariance of the estimated odds ratio for the logistic model is a logical result. In Section 2.5.1, we showed that the assumption of equal misclassification
probabilities of covariates over all exposure categories (i.e., \( \phi_{sr} = \phi_{sri} \) for all \( i = 0, 1, \ldots, k \)) is equivalent to the assumption that, given classified covariate levels, exposure and true covariate levels are independent. Consider the case of logistic regression. This assumption implies that, if the classified covariates are being controlled for, controlling for the true covariates in addition will not affect the degree of confounding. In other words, adjusting for the classified covariates will result in the same adjusted odds ratio as adjusting for the true covariates.

The model used to control for the classified covariates is

\[
\text{logit } \theta = \beta_a + \beta_1 + \Phi^\alpha V \gamma .
\]

The model used to control for the true covariates is

\[
\text{logit } \theta = \beta_a + \beta_1 + \Phi V \gamma .
\]

The above argument implies that setting \( \phi_{sr} = \phi_{sri} \) for all \( i = 0, 1, \ldots, k \) will result in equal \( \beta_1 \)'s for the two models. This agrees with the result presented earlier in this section, namely that \( \hat{\beta} \) is invariant to choices of \( \Phi \) assuming that \( \phi_{sr} = \phi_{sri} \) for all \( i \).

4.3 Nonconstant Misclassification Probabilities

In developing the theory in Chapter 3, one basic assumption was that misclassification probabilities were constant over all strata,

\[
i.e., \Pi = \begin{bmatrix} \Pi_1 \\ \vdots \\ \Pi_k \end{bmatrix} . \quad \text{This assumption was necessary to show the}
\]

the invariance of \( \hat{R} \) in the LS case, which led to the invariance of
\( \hat{\gamma}_{1s}, \hat{\Pi}_{1s} \) and \( \hat{\gamma}_{1s} \). In the WLS case, the matrix \( \hat{R}_w \) was found to be invariant under this assumption, which led to the invariance of several WLS estimators. This invariance, in turn, contributed to the invariance of several ML estimators. Now, we will investigate the effect of a \( \hat{\Pi} \) which does not meet this assumption.

For simplicity's sake, consider a situation in which there are just two strata, i.e., \( S=2 \). Let \( \hat{\Pi} = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{bmatrix} \) where \( \hat{\Pi}_1 \) is the matrix of estimated misclassification probabilities corresponding to the first stratum and \( \hat{\Pi}_2 \) is the matrix of estimated misclassification probabilities corresponding to the second stratum. Using this definition of \( \hat{\Pi} \), let us expand the matrix expression

\[
\hat{\Pi}(\hat{\Pi}'\hat{\Pi})^{-1}\hat{\Pi}' = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{bmatrix} \left( (\hat{\Pi}_1, \hat{\Pi}_2)' \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{bmatrix} \right)^{-1} (\hat{\Pi}_1, \hat{\Pi}_2)'
\]

\[
= \begin{bmatrix} (\hat{\Pi}_1^{-1})^{-1} \\ (\hat{\Pi}_2^{-1})^{-1} \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1, \hat{\Pi}_1 + \hat{\Pi}_2 \hat{\Pi}_2 \end{bmatrix}^{-1} (\hat{\Pi}_1, \hat{\Pi}_2)'
\]

\[
= \begin{bmatrix} (\hat{\Pi}_1, \hat{\Pi}_2)^{-1} \\ (\hat{\Pi}_1, \hat{\Pi}_2)^{-1} \end{bmatrix} \begin{bmatrix} (\hat{\Pi}_1^{-1}), (\hat{\Pi}_2^{-1})^{-1} \\ (\hat{\Pi}_1^{-1}), (\hat{\Pi}_2^{-1})^{-1} \end{bmatrix}
\]
\[
\begin{bmatrix}
\hat{\Pi}_1^{-1}\hat{\Pi}_1' + \hat{\Pi}_1'^{-1}\hat{\Pi}_2'\hat{\Pi}_2\hat{\Pi}_1^{-1}
\hat{\Pi}_2'^{-1}\hat{\Pi}_1' + \hat{\Pi}_2'^{-1}\hat{\Pi}_2\hat{\Pi}_1^{-1}
\end{bmatrix}^{-1}
\begin{bmatrix}
\hat{\Pi}_1^{-1}\hat{\Pi}_1' + \hat{\Pi}_1'^{-1}\hat{\Pi}_2'\hat{\Pi}_2\hat{\Pi}_1^{-1}
\hat{\Pi}_2'^{-1}\hat{\Pi}_1' + \hat{\Pi}_2'^{-1}\hat{\Pi}_2\hat{\Pi}_1^{-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\hat{\Pi}_1^{-1}\hat{\Pi}_1' + \hat{\Pi}_1'^{-1}\hat{\Pi}_2'\hat{\Pi}_2\hat{\Pi}_1^{-1}
\hat{\Pi}_2'^{-1}\hat{\Pi}_1' + \hat{\Pi}_2'^{-1}\hat{\Pi}_2\hat{\Pi}_1^{-1}
\end{bmatrix}^{-1}
\begin{bmatrix}
\hat{\Pi}_1^{-1}\hat{\Pi}_1' + \hat{\Pi}_1'^{-1}\hat{\Pi}_2'\hat{\Pi}_2\hat{\Pi}_1^{-1}
\hat{\Pi}_2'^{-1}\hat{\Pi}_1' + \hat{\Pi}_2'^{-1}\hat{\Pi}_2\hat{\Pi}_1^{-1}
\end{bmatrix}
\]

Unlike the situation in which \( \hat{\Pi}_1 = \hat{\Pi}_2 \), the matrix \( \hat{\Pi} (\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi}' \) will not be invariant to choices of \( \hat{\Pi}_1 \) and \( \hat{\Pi}_2 \) when \( \hat{\Pi}_1 \neq \hat{\Pi}_2 \). From this, it follows that the matrix \( \hat{R} \) will not be invariant to choices of \( \hat{\Pi}_1 \) and \( \hat{\Pi}_2 \). This result is easily expanded to any number of strata. In general, if there are \( S \) strata, then

\[
\hat{\Pi} = \begin{bmatrix}
\hat{\Pi}_1 \\
\hat{\Pi}_2 \\
\vdots \\
\hat{\Pi}_S
\end{bmatrix}, \text{ and the matrix expression } \hat{\Pi} (\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi}', \text{ and}
\]

therefore \( \hat{R} \), will not be invariant to choices of \( \hat{\Pi} \).

Similarly, if \( \hat{\Pi} = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{bmatrix} \) the matrix

\[
\hat{\Pi} (\hat{\Pi}' D_x \hat{\Pi})^{-1} \hat{\Pi}' = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_2 \end{bmatrix} \begin{bmatrix} \hat{\Pi}_1'^{-1} \hat{\Pi}_1' + \hat{\Pi}_2'^{-1} \hat{\Pi}_2 D_x \hat{\Pi}_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}_1', \hat{\Pi}_2' \end{bmatrix}
\]

is also not invariant to choices of \( \hat{\Pi}_1 \) and \( \hat{\Pi}_2 \). This result clearly
holds for any number of strata, which implies that the matrix \( \hat{R}_w \) is not invariant for this situation.

Not only are \( \hat{R} \) and \( \hat{R}_w \) not invariant, but since the LS, WLS and ML estimates of \( y \) are dependent on these two matrices, they will also not be invariant to choices of \( \Pi \). However, if \( \hat{\Pi}=\Pi \), it is easily shown that \( \hat{\beta}_{ls} \) and \( \hat{\gamma}_{ls} \) are unbiased estimators of \( \beta^* \) and \( \gamma^* \), and that the WLS and ML estimators of \( \beta^* \) and \( \gamma^* \) are asymptotically unbiased.

Similar results are found for model (4.2). The invariance of \( \hat{\beta}_{ls}, \hat{\beta}_{wls}, \) and \( \hat{\beta}_{ml} \) to changes in \( \hat{\phi} \) does not hold if \( \phi_{sr} \neq \phi_{sri} \) for all \( i=0,1,\ldots,k \). In this case, there still exists a matrix \( T \) such that \( \hat{\phi}\hat{V}=\hat{V}T \); however, this matrix \( T \) is not invertible since it is not a square matrix. Without the invertibility of \( T \), the invariance of \( \hat{\beta} \) does not hold.

These results indicate that the equality of exposure misclassification probabilities over strata or covariate misclassification probabilities over exposure categories can have a great effect on the properties of certain estimators. In the case of misclassification of covariates, in fact, if one can ascertain that the misclassification probabilities are equal over all exposure categories, the misclassification error need not be corrected for. Therefore, determining whether the misclassification probabilities are constant may be one of the most important steps in correcting for misclassification error.
4.4 Ordinal Exposure Categories

A short, but important, note is made here concerning situations in which the exposure categories are ordinal, rather than nominal. In these cases, models such as those developed in Section 2.7 can be applied when there is misclassification of exposure. The \( \Pi \) matrix associated with those models, however, lacks certain properties which the \( \Pi \) matrices of models (1.16), (1.18), and (1.19) possess.

As was illustrated in Section 2.7, the columns of the \( \Pi \) matrix for model (2.10) are not composed of the misclassification probabilities themselves, as has been the case with the other \( \Pi \) matrices we have considered. Instead, the elements are weighted averages of the scores, \( d_j \) for \( j=0,1,\ldots,k \), with the weights being the misclassification probabilities. Likewise, the elements of the \( \Pi \) matrix for model (2.11) are weighted averages of the scores and of the scores taken to certain powers.

Although these \( \Pi \) matrices can be written as

\[
\Pi = \begin{bmatrix}
\Pi_1 \\
\Pi_2 \\
\vdots \\
\Pi_k
\end{bmatrix},
\]

the individual \( \Pi_i \) matrices are not invertible. Neither are they the product of two matrices, of which one is invertible and the other invariant, as was the matrix corresponding to the misclassified variable in the previous section.

The consequence of this is that the statistics found earlier to be invariant to choices of misclassification probabilities are not invariant in this case. The values of \( \hat{\gamma} \), \( \text{Var}(\hat{\gamma}) \), and the goodness-of-fit statistics will vary with different sets of \( \pi_{ij} \)'s for these models.
4.5 Relationship Between $\hat{\beta}^*$ and $\hat{\beta}^0$

In this section, a relationship is derived between estimates of $\beta$ obtained when $\hat{\Pi} = \Pi$ and those obtained when $\hat{\Pi} = \hat{\Pi}^0$. Let us define $\hat{\beta}^0 = (\hat{\beta}_{\alpha}^0, \hat{\beta}_1^0, \ldots, \hat{\beta}_k^0)'$ as the vector of estimates of $\beta$ obtained when misclassification is ignored (i.e., when $\hat{\Pi} = \hat{\Pi}^0$). $\hat{\beta}^* = (\hat{\beta}_{\alpha}^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_k^*)'$ is defined as the vector of estimates obtained when $\hat{\Pi} = \Pi$. The relationship found between $\hat{\beta}^0$ and $\hat{\beta}^*$ for LS, WLS, and ML estimation methods is used in later sections to derive several results concerning the bias of $\hat{\beta}^0$ and properties of certain hypothesis tests.

Let us begin by examining the two $\hat{\Pi}$ matrices involved, $\Pi$ and $\hat{\Pi}^0$. Recall from Section 1.10 that $\Pi$ can be written as

$$\Pi = \begin{bmatrix} \Pi_1 \\ \Pi_1 \\ \vdots \\ \Pi_1 \end{bmatrix} \quad \text{where} \quad \Pi_1 = \begin{bmatrix} 1 & \pi_{01} & \cdots & \pi_{0k} \\ 1 & \pi_{11} & \cdots & \pi_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \pi_{1k} & \cdots & \pi_{kk} \end{bmatrix}.$$ 

Similarly, $\hat{\Pi}^0$ can be written as

$$\hat{\Pi}^0 = \begin{bmatrix} \hat{\Pi}^0_1 \\ \hat{\Pi}^0_1 \\ \vdots \\ \hat{\Pi}^0_1 \end{bmatrix} \quad \text{where} \quad \hat{\Pi}^0_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$ 

The elements in the first column and main diagonal of $\hat{\Pi}^0_1$ are all 1's; the remaining elements are all 0's.

$\Pi_1$ and $\hat{\Pi}^0_1$ are related through the following expression:
\[ \Pi_1 = \hat{\Pi}_1^0 \Delta \]  

where

\[
\Delta = \begin{bmatrix}
1 & \pi_{11} & \cdots & \pi_{0k} \\
0 & (\pi_{11} - \pi_{01}) & \cdots & (\pi_{1k} - \pi_{0k}) \\
\vdots & \vdots & \ddots & \vdots \\
0 & (\pi_{k1} - \pi_{01}) & \cdots & (\pi_{kk} - \pi_{0k})
\end{bmatrix}
\]

In the case of LS estimation, we can easily derive the relationship between \( \hat{\beta}^*_{ls} \) and \( \hat{\beta}^0_{ls} \) using expression (4.5). From expression (3.5), the two vectors of estimates can be written

\[
\hat{\beta}^*_{ls} = (\Pi' \hat{\Pi})^{-1} \Pi' (Y - \hat{V} \hat{\gamma}^*_{ls})
\]

and

\[
\hat{\beta}^0_{ls} = (\hat{\Pi}^0' \hat{\Pi}^0)^{-1} \hat{\Pi}^0' (Y - \hat{V} \hat{\gamma}^0_{ls})
\]

where \( \hat{\gamma}^*_{ls} \) and \( \hat{\gamma}^0_{ls} \) are the LS estimates of \( \gamma \) obtained when \( \Pi \) and \( \hat{\Pi}^0 \) are used in the fitted model, respectively.

Since \( \Pi_1 \) and \( \hat{\Pi}_1^0 \) are both invertible, we can see that

\[
(\Pi_1' \Pi_1)^{-1} \Pi_1' = \Pi_1^{-1}
\]

and

\[
(\hat{\Pi}^0_1' \hat{\Pi}^0_1)^{-1} \hat{\Pi}^0_1' = \hat{\Pi}^0_1^{-1}
\]

Using this fact, along with expression (3.9), we find

\[
\hat{\beta}^*_{ls} = \frac{1}{S} \left[ \Pi_1^{-1}, \Pi_1^{-1}, \ldots, \Pi_1^{-1} \right] \left[ Y - \hat{V} \hat{\gamma}^*_{ls} \right]
\]

and

\[
\hat{\beta}^0_{ls} = \frac{1}{S} \left[ \hat{\Pi}_1^0, \hat{\Pi}_1^0, \ldots, \hat{\Pi}_1^0 \right] \left[ Y - \hat{V} \hat{\gamma}^0_{ls} \right] .
\]

Now, since \( \Delta \) is invertible, expression (4.5) implies

\[
\Pi_1^{-1} = \Delta^{-1} \hat{\Pi}_1^0
\]

Substituting expression (4.7) into expression (4.6) gives us
\[ \hat{\beta}^*_{ls} = \frac{1}{S} \left[ \Delta^{-1} \hat{\pi}_0^{-1}, \Delta^{-1} \hat{\pi}_1^{-1}, \ldots, \Delta^{-1} \hat{\pi}_1^{-1} \right] \left[ Y - V \hat{\gamma}^*_{ls} \right] \]

\[ = \frac{1}{S} \Delta^{-1} \left[ \hat{\pi}_0^{-1}, \hat{\pi}_1^{-1}, \ldots, \hat{\pi}_1^{-1} \right] \left[ Y - V \hat{\gamma}^*_{ls} \right]. \quad (4.8) \]

It was shown in Chapter 3 that the vector \( \hat{\gamma}_{ls} \) is invariant to choices of \( \hat{\Pi} \). This implies that \( \hat{\gamma}^*_{ls} = \hat{\gamma}^0_{ls} \). Thus, expression (4.8) becomes

\[ \hat{\beta}^*_{ls} = \frac{1}{S} \Delta^{-1} \left[ \hat{\pi}_0^{-1}, \hat{\pi}_1^{-1}, \ldots, \hat{\pi}_1^{-1} \right] \left[ Y - V \hat{\gamma}^0_{ls} \right] \]

\[ = \Delta^{-1} \hat{\beta}^0_{ls}, \]

which implies \( \hat{\beta}^0_{ls} = \Delta \hat{\beta}^*_{ls} \). \( \quad (4.9) \)

Consider a situation in which there are just two exposure categories. In this case,

\[ \hat{\Pi}_1 = \begin{bmatrix} 1 & \pi_{01} \\ 1 & \pi_{11} \end{bmatrix}, \quad \hat{\Pi}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \Delta = \begin{bmatrix} 1 & \pi_{01} \\ 0 & (\pi_{11} - \pi_{01}) \end{bmatrix}. \]

Using expression (4.9), we find

\[ \begin{bmatrix} \hat{\beta}^0_\alpha \\ \hat{\beta}^0_1 \end{bmatrix} = \begin{bmatrix} 1 & \pi_{01} \\ 0 & (\pi_{11} - \pi_{01}) \end{bmatrix} \begin{bmatrix} \hat{\beta}^*_\alpha \\ \hat{\beta}^*_1 \end{bmatrix} \]

\[ \Rightarrow \begin{bmatrix} \hat{\beta}^0_\alpha \\ \hat{\beta}^0_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}^*_{\alpha} + \pi_{01} \hat{\beta}^*_1 \\ (\pi_{11} - \pi_{01}) \hat{\beta}^*_1 \end{bmatrix} \]

\[ \Rightarrow \quad \text{E}(\hat{\beta}^0_\alpha | \hat{\Pi}) = \beta^*_{\alpha} + \pi_{01} \beta^*_1 \quad \text{and} \quad \text{E}(\hat{\beta}^0_1 | \hat{\Pi}) = (\pi_{11} - \pi_{01}) \beta^*_1. \]

These are consistent with the results found in Section 3.2.2 for
the bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ when any $\Pi$ is used in the fitted model. Substituting $\hat{\pi}_{11}=1$ and $\hat{\pi}_{01}=0$ into the expression given in that section leads to the expression given above.

The relationship in expression (4.9) will be shown to hold for both WLS and ML estimates. Beginning with the WLS case, $\hat{\beta}_{\text{wls}}^*$ and $\hat{\beta}_{\text{wls}}^0$ can be written using expression (3.18) as

$$\hat{\beta}_{\text{wls}}^* = (\Pi'D_x\Pi)^{-1}\Pi'D_x(Y-\hat{V}_w\hat{y}_{\text{wls}})$$

and

$$\hat{\beta}_{\text{wls}}^0 = (\hat{\Pi}'D_x\hat{\Pi})^{-1}\hat{\Pi}'D_x(Y-\hat{V}_w\hat{y}_{\text{wls}}).$$

The matrix $D_x$, a function of the observed counts, is constant for all values of $\Pi$, including the two choices used above, $\Pi$ and $\hat{\Pi}^0$. Since $\hat{\gamma}_{\text{wls}}$ is invariant to choices of $\Pi$, we have substituted $\hat{\gamma}_{\text{wls}} = \hat{\gamma}_{\text{wls}}^*$ and $\hat{\gamma}_{\text{wls}} = \hat{\gamma}_{\text{wls}}^0$ into the above expressions.

We showed in Section 3.3.1 that

$$(\Pi'D_x\Pi)^{-1} = \Pi_i^{-1}\hat{\Pi}_i^{-1}\hat{\Pi}_i^{-1}.$$ This expression holds for $\hat{\Pi} = \hat{\Pi}^0$ and $\hat{\Pi} = \Pi$, so that

$$(\Pi^0'D_x\Pi^0)^{-1} = \Pi_1^{-1}\hat{\Pi}^0_1^{-1}\hat{\Pi}^0_1^{-1} \text{ and } (\Pi'D_x\Pi)^{-1} = \Pi_i^{-1}\hat{\Pi}_i^{-1}\hat{\Pi}_i^{-1}.$$ This implies that

$$\hat{\beta}_{\text{wls}}^* = \Pi_1^{-1}\hat{\Pi}^{-1}_1\Pi_1^{-1}\hat{\Pi}_1^{-1} \left[ \Pi_1',\Pi_1',\ldots,\Pi_1' \right] D_x(Y-\hat{V}_w\hat{y}_{\text{wls}})$$

$$= \Pi_1^{-1}\hat{\Pi}_1^{-1} \left[ I_{k+1},I_{k+1},\ldots,I_{k+1} \right] D_x(Y-\hat{V}_w\hat{y}_{\text{wls}}) \quad (4.10)$$

Similarly,

$$\hat{\beta}_{\text{wls}}^0 = \hat{\Pi}_1^{-1}\hat{\Pi}_1^{-1} \left[ I_{k+1},I_{k+1},\ldots,I_{k+1} \right] D_x(Y-\hat{V}_w\hat{y}_{\text{wls}}).$$

Substituting expression (4.7) into expression (4.10) leads us to
\[ \hat{\beta}^{*}_{\text{wls}} = \Delta^{-1} \Pi_{0}^{-1} \hat{\Gamma} \left( I_{k+1}, I_{k+1}, \ldots, I_{k+1} \right) D_{X}(Y-V_{y_{\text{wls}}}) \]
\[ = \Delta^{-1} \hat{\beta}^{0}_{\text{wls}}, \]

which implies

\[ \hat{\beta}^{0}_{\text{wls}} = \Delta \hat{\beta}^{*}_{\text{wls}}. \]

Similarly, this relationship can be shown to hold for \( \hat{\beta}^{*}_{\text{ml}} \) and \( \hat{\beta}^{0}_{\text{ml}} \) by examining the subsequent-step formulas for the two vectors of estimators. In this case, however, we will show the relationship holds using a simpler proof which could be applied to any of the three types of estimators. This proof makes use of the invariance of \( \hat{\Pi}_{0} \beta^{0}_{\text{ml}} \) for any choice of \( \hat{\Pi} \), which implies

\[ \hat{\Pi}^{0} \hat{\beta}^{0}_{\text{ml}} = \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \]

\[ \Rightarrow \begin{bmatrix} \hat{\Pi}^{0} \hat{\beta}^{0}_{\text{ml}} \\ \hat{\Pi}^{0} \hat{\beta}^{0}_{\text{ml}} \\ \vdots \\ \hat{\Pi}^{0} \hat{\beta}^{0}_{\text{ml}} \end{bmatrix} = \begin{bmatrix} \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \\ \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \\ \vdots \\ \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \end{bmatrix} \]

\[ \Rightarrow \hat{\Pi}^{0} \hat{\beta}^{0}_{\text{ml}} = \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \]

\[ \Rightarrow \hat{\beta}^{0}_{\text{ml}} = \hat{\Pi}^{0}^{-1} \hat{\Pi} \hat{\beta}^{*}_{\text{ml}} \quad (4.11) \]

From expression (4.5) we can see that

\[ \Delta = \hat{\Pi}^{0}^{-1} \hat{\Pi} \]

Substituting this expression into expression (4.11) leads to

\[ \hat{\beta}^{0}_{\text{ml}} = \Delta \hat{\beta}^{*}_{\text{ml}} \]

which is the desired result.
Therefore, we have seen that the above relationship holds for all three types of estimators. The relationship also holds for the corresponding parameters. We can see this through the LS estimators. When \( \hat{\Pi} = \hat{\Pi}^0 \), expression (3.12) for the expected value of \( \hat{\beta}_{ls} \) becomes

\[
E(\hat{\beta}^{0}_{ls} | \hat{\Pi} = \hat{\Pi}^0 ) = \beta^0 = \hat{\Pi}^{0\top} \Pi \beta^*
\]

which implies \( \beta^0 = \Delta \beta^* \).

4.5 Bias Towards The Null Of \( \hat{\beta}^{0}_{-k} \)

In this section, we will prove a result similar to that of Gladen and Rogan (1979) discussed in Section 1.6.2. They found bias towards the null in estimating the relative risk which compares the highest and lowest exposure categories when misclassification is ignored. They also showed this property does not necessarily hold for the relative risks comparing intermediate categories to the lowest.

Here, we will show that \( \hat{\beta}^{0}_{-k} \) is biased towards the null value of 0. \( \hat{\beta}^{0}_{-k} \) is the estimate which is used to compare the highest and lowest exposure categories when misclassification is ignored. In the case of logistic regression, this result implies that the estimate of \( OR_k \) is biased towards the null when misclassification is ignored.

Expanding the relationship found for the parameters, namely \( \beta^0 = \Delta \beta^* \), we see
\[
\begin{pmatrix}
\beta^0_0 \\
\beta^0_1 \\
\vdots \\
\beta^0_k
\end{pmatrix} =
\begin{pmatrix}
1 & \pi_{01} & \cdots & \pi_{0k} \\
0 & (\pi_{11} - \pi_{01}) & \cdots & (\pi_{1k} - \pi_{0k}) \\
\vdots & \vdots & \ddots & \vdots \\
0 & (\pi_{k1} - \pi_{01}) & \cdots & (\pi_{kk} - \pi_{0k})
\end{pmatrix}
\begin{pmatrix}
\beta^*_0 \\
\beta^*_1 \\
\vdots \\
\beta^*_k
\end{pmatrix}
\]
which implies
\[
\beta^0_0 = \beta^*_0 + \sum_{j=1}^{k} \pi_{0j} \beta^*_j,
\]
\[
\beta^0_1 = \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \beta^*_j,
\]
\[
\vdots
\]
\[
\beta^0_k = \sum_{j=1}^{k} (\pi_{kj} - \pi_{0j}) \beta^*_j.
\]

To show bias towards the null of \(\hat{\beta}^0_k\), we need to show that
\[
E(\hat{\beta}^0_k) = \beta^0_k < \beta^*_k = E(\hat{\beta}^*_k).
\]
To do this, we will make an assumption equivalent to one made by Gladen and Rogan, namely that \(\theta_0^* < \theta_1^* < \cdots < \theta_k^*\). This assumption implies that \(0 < \beta_1^* < \beta_2^* < \cdots < \beta_k^*\).

As we saw above,
\[
\beta^0_k = \sum_{j=1}^{k} (\pi_{kj} - \pi_{0j}) \beta^*_j
\]
\[
= \sum_{j \in \omega} (\pi_{kj} - \pi_{0j}) \beta^*_j + \sum_{j \in \Omega} (\pi_{kj} - \pi_{0j}) \beta^*_j
\]
where \(\omega = \{ j > 0 : (\pi_{kj} - \pi_{0j}) \leq 0 \}\)
and \(\Omega = \{ j > 0 : (\pi_{kj} - \pi_{0j}) > 0 \}\)
\[
\Rightarrow \beta^0_k \leq \sum_{j \in \Omega} (\pi_{kj} - \pi_{0j}) \beta^*_j
\]
\[ < \sum_{j \in \Omega} (\pi_{kj} - \pi_{0j}) \beta^*_k \]

\[ = \left[ \pi_{kk} - \pi_{0k} + \sum_{j \in \Omega} \pi_{kj} - \sum_{j \in \Omega} \pi_{0j} \right] \beta^*_k \quad (4.12) \]

since \( j = k \in \Omega \).

Now,

\[ 0 < \left[ \sum_{j \in \Omega} \pi_{kj} - \sum_{j \in \Omega} \pi_{0j} \right] \leq (1 - \pi_{kk}) \]

since the maximum value of \( \sum_{j \in \Omega} \pi_{kj} \) is \( (1 - \pi_{kk}) \), the minimum value \( \sum_{j \in \Omega, j \neq k} \pi_{kj} \) of \( \sum_{j \in \Omega} \pi_{0j} \) is 0, and \( j \in \Omega \) ensures that \( \pi_{kj} > \pi_{0j} \).

Using this inequality, we can see that

\[ \left[ \pi_{kk} - \pi_{0k} + \sum_{j \in \Omega} \pi_{kj} - \sum_{j \in \Omega} \pi_{0j} \right] \leq (\pi_{kk} - \pi_{0k}) + (1 - \pi_{kk}) = 1 - \pi_{0k} \leq 1 \]

and

\[ \left[ \pi_{kk} - \pi_{0k} + \sum_{j \in \Omega} \pi_{kj} - \sum_{j \in \Omega} \pi_{0j} \right] > \pi_{kk} - \pi_{0k} > 0 \]

This shows us that

\[ 0 < \left[ \pi_{kk} - \pi_{0k} + \sum_{j \in \Omega} \pi_{kj} - \sum_{j \in \Omega} \pi_{0j} \right] \leq 1 \]

which can be used, along with expression (4.12), to show that
\[ \beta^0_k < \left[ \pi_{kk} - \pi_{0k} + \sum_{j \in \Omega} \pi_{kj} - \sum_{j \notin k} \pi_{oj} \right] \beta^*_k \leq \beta^*_k \]

or simply,
\[ \beta^0_k < \beta^*_k. \]

Therefore, we have shown that \( \hat{\beta}^0_k \) is biased towards the null. This implies that \( \hat{OR}^0_k \), the estimated odds ratio comparing the highest and lowest exposure categories obtained when misclassification is ignored, is biased towards the null also.

Nothing can be said, however, concerning the direction of the bias for estimated odds ratios involving intermediate exposure categories.

4.7 Relationship Between Variances

In this section, we will use the relationship \( \hat{\beta}^0 = \Delta \hat{\beta}^* \), which was found to hold for LS, WLS, and ML estimators, to derive a relationship between the variances of the \( \hat{\beta}^0 \)'s and the \( \hat{\beta}^* \)'s.

We saw in Chapter 3 that \( \text{Var}(\hat{\beta}) \) varies with choices of \( \Pi \) for LS, WLS, and ML estimation. Therefore, \( \text{Var}(\hat{\beta}^0) \) and \( \text{Var}(\hat{\beta}) \) will certainly not be equal to each other regardless of the estimation procedure used.

\[ \hat{\beta}^0 = \Delta \hat{\beta}^* \text{ implies } \text{Var}(\hat{\beta}^0) = \Delta \text{Var}(\hat{\beta}^*) \Delta'. \]  

(4.13)

Let \( \sigma^2_{ij} = \text{Cov}(\hat{\beta}_i, \hat{\beta}_j) \) for \( i,j=1,2,\ldots,k \),
\[ \sigma^2_{0j} = \text{Cov}(\hat{\beta}_0, \hat{\beta}_j) \text{ for } j=1,2,\ldots,k, \text{ and} \]
\[ \sigma^2_{00} = \text{Cov}(\hat{\beta}_0, \hat{\beta}_0) \text{ so that} \]
\[ \text{Var}(\hat{\beta}) = \begin{bmatrix} \sigma^2_{00} & \sigma^2_{01} & \cdots & \sigma^2_{0k} \\ \sigma^2_{10} & \sigma^2_{11} & \cdots & \sigma^2_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{k0} & \sigma^2_{k1} & \cdots & \sigma^2_{kk} \end{bmatrix} \]

and

\[ \text{Var}(\hat{\beta}^0) = \begin{bmatrix} \sigma^2_{00} & \sigma^2_{01} & \cdots & \sigma^2_{0k} \\ \sigma^2_{10} & \sigma^2_{11} & \cdots & \sigma^2_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{k0} & \sigma^2_{k1} & \cdots & \sigma^2_{kk} \end{bmatrix} \]

then \( \Delta \text{Var}(\hat{\beta}) = \)

\[ \begin{bmatrix} \sigma^2_{00} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{0j} & \sigma^2_{01} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{1j} & \cdots & \sigma^2_{0k} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{kj} \\ \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \sigma^2_{0j} & \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \sigma^2_{1j} & \cdots & \sum_{j=1}^{k} (\pi_{1j} - \pi_{0j}) \sigma^2_{kj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{k} (\pi_{kj} - \pi_{0j}) \sigma^2_{0j} & \sum_{j=1}^{k} (\pi_{kj} - \pi_{0j}) \sigma^2_{1j} & \cdots & \sum_{j=1}^{k} (\pi_{kj} - \pi_{0j}) \sigma^2_{kj} \end{bmatrix} \]

Looking at just the diagonal elements of \( \text{Var}(\hat{\beta}^0) \), which are the variances of \( \hat{\beta}^0_\alpha \) and the \( \hat{\beta}^0_j \)'s, we can see that expression (4.13) leads to

\[ \text{Var}(\hat{\beta}^0) = \Delta \text{Var}(\hat{\beta}) \triangle \Rightarrow \]

111
\[
\sigma^2_{00} = \sigma^2_{00} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{0j} + \pi_{01}(\sigma^2_{01} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{ij}) \\
+ \ldots + \pi_{0k}(\sigma^2_{0k} + \sum_{j=1}^{k} \pi_{0j} \sigma^2_{kj})
\]

\[
= \sigma^2_{00} + 2\sum_{j=1}^{k} \pi_{0j} \sigma^2_{0j} + \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_{0i} \pi_{0j} \sigma^2_{ij}
\]

\[
\sigma^2_{11} = (\pi_{11}-\pi_{01})\sum_{j=1}^{k} (\pi_{1j}-\pi_{0j}) \sigma^2_{1j} + (\pi_{12}-\pi_{02})\sum_{j=1}^{k} (\pi_{1j}-\pi_{0j}) \sigma^2_{2j} \\
+ \ldots + (\pi_{1k}-\pi_{0k})\sum_{j=1}^{k} (\pi_{1j}-\pi_{0j}) \sigma^2_{kj}
\]

\[
= \sum_{i=1}^{k} \sum_{j=1}^{k} (\pi_{1i}-\pi_{0i})(\pi_{1j}-\pi_{0j}) \sigma^2_{kj}
\]

\[
\sigma^2_{kk} = (\pi_{k1}-\pi_{01})\sum_{j=1}^{k} (\pi_{kj}-\pi_{0j}) \sigma^2_{1j} + (\pi_{k2}-\pi_{02})\sum_{j=1}^{k} (\pi_{kj}-\pi_{0j}) \sigma^2_{2j} \\
+ \ldots + (\pi_{kk}-\pi_{0k})\sum_{j=1}^{k} (\pi_{kj}-\pi_{0j}) \sigma^2_{kj}
\]

\[
= \sum_{i=1}^{k} \sum_{j=1}^{k} (\pi_{ki}-\pi_{0i})(\pi_{kj}-\pi_{0j}) \sigma^2_{ij}
\]

Since \((\pi_{ij}-\pi_{0j})\) and \(\sigma^2_{ij}\) for \(i\neq j\) can be either positive or negative for \(i=0,1,\ldots,k\) and \(j=0,1,\ldots,k\), we can make no conclusions concerning the direction of the relationship between \(\sigma^2_{0j}\) and \(\sigma^2_{jj}\) for \(j=0,1,\ldots,k\).

The relationship between the true variances, given in expression (4.13), also holds for the estimated asymptotic variances for the ML estimators of \(\beta\). From the formula given in Section 3.4.2 and
its expansion in Section 3.4.1, the estimated asymptotic variance of \( \hat{\beta}_{ml} \) can be written as

\[
\text{Var}_a(\hat{\beta}_{ml}) = \Pi_1^{-1} P \Pi_1^{-1}'
\]  

(4.14)

where the matrix \( P \) is a matrix whose elements are functions of \( \hat{\mu}_{ml} \). As we saw in Chapter 3, \( \hat{\mu}_{ml} \) is invariant to choices of \( \hat{\Pi} \). Therefore, \( P \) is also invariant.

The estimated asymptotic variance of \( \hat{\beta}_{ml} \) can be written as

\[
\text{Var}_a(\hat{\beta}_{ml}) = \hat{\Pi}_1^{-1} P \hat{\Pi}_1^{-1}'
\]

The matrix \( P \) in this formula is identical to the matrix \( P \) in formula (4.14) since it is invariant to the choice of \( \hat{\Pi} \). Now, substituting expression (4.7) into (4.14) leads to

\[
\text{Var}_a(\hat{\beta}_{ml}) = \Delta^{-1} \hat{\Pi}_1^{-1} P \hat{\Pi}_1^{-1} \Delta'
\]

\[= \Delta^{-1} \text{Var}_a(\hat{\beta}_{ml}) \Delta'
\]

which implies

\[
\text{Var}_a(\hat{\beta}_{ml}) = \Delta \text{Var}_a(\hat{\beta}_{ml}) \Delta'.
\]

Thus, the relationship between the individual true asymptotic variances of the \( \hat{\beta}_{j} \)'s and \( \hat{\beta}_{j} \)'s, shown on the previous page, will also hold for their estimated asymptotic variances.

4.8 Invariance Of Test Of \( H_0: \beta_1=\beta_2=\ldots=\beta_k=0 \)

The null hypothesis \( H_0: \beta_1=\beta_2=\ldots=\beta_k=0 \) is equivalent, in the context of logistic regression, to the hypothesis \( H_0: \text{OR}_1=\text{OR}_2=\ldots=\text{OR}_k=1 \). In general, it is a hypothesis which says that there is no exposure-disease relationship.

Suppose a contingency table is formed whose rows are defined
by levels of exposure and whose columns are defined by the presence or absence of disease. Further, suppose we are interested in testing the hypothesis that the exposure and disease variables of the contingency table are independent. We saw in Section 1.6 that several authors have found that the Type I error rate for such a test of independence is not affected by nondifferential misclassification. We will show that this property holds in the context of hypothesis testing using ML estimation. Specifically, we will show that a test of $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$ produces a p-value which is invariant to choices of $\Pi$.

The Wald statistic used for significance testing of contrast statements in CATMAX was discussed in Section 2.3. In order to test $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$, the statistic becomes

$$Q_{wc} = \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right)' C' \left( C \text{Var}_{a} \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right) C' \right)^{-1} C \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right).$$

The matrix $C$ can be expressed as the horizontal concatenation of two submatrices, $C_\beta$ and $C_\gamma$, i.e., $C = (C_\beta, C_\gamma)$, where $C_\beta$ is that part of $C$ which pertains to $\beta$ and $C_\gamma$ is that part of $C$ which pertains to $\gamma$. Since $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$ makes no stipulations concerning the value of $\gamma$, $C_\gamma = 0$. Therefore, $C = (C_\beta, 0)$ and

$$C \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right) = C_\beta \hat{\beta}_{ml}. \text{ Also, } C \text{Var}_{a} \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right) C' = C_\beta C_\beta'. \text{ Therefore, } C = (C_\beta, 0) \text{ and}$$

$$C \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right) = C_\beta \hat{\beta}_{ml}. \text{ Also, } C \text{Var}_{a} \left( \hat{\beta}_{ml}, \hat{\gamma}_{ml} \right) C' = C_\beta \text{Var}_{a} \left( \hat{\beta}_{ml} \right) C_\beta'.$$

Now we can write $Q_{wc}$ as

$$Q_{wc} = \beta_{ml}' C_\beta' \left( C_\beta \text{Var}_{a} \left( \hat{\beta}_{ml} \right) C_\beta' \right)^{-1} C_\beta \beta_{ml}.$$
where
\[
\mathbf{C}_\beta = \begin{bmatrix}
0 & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 0 & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{bmatrix}
\]

If \( \hat{\Pi} = \hat{\Pi}^0 \) is the matrix of misclassification probabilities used in the fitted model, \( Q_{wc} \) becomes
\[
\begin{align*}
Q_{wc}^0 &= \hat{\beta}_{ml}^* \mathbf{C}_\beta' \left( \mathbf{C}_\beta \mathbf{Var}_a(\hat{\beta}_{ml}^0) \mathbf{C}_\beta \right)^{-1} \mathbf{C}_\beta \hat{\beta}_{ml}^0 \\
&= \hat{\beta}_{ml}^* \left( \mathbf{C}_\beta \mathbf{Var}_a(\hat{\beta}_{ml}^*\Delta') \mathbf{C}_\beta \right)^{-1} \mathbf{C}_\beta \Delta \hat{\beta}_{ml}^* \\
&= \hat{\beta}_{ml}^* \left( \mathbf{C}_\beta \mathbf{Var}_a(\hat{\beta}_{ml}^*\Delta') \mathbf{C}_\beta \right)^{-1} \mathbf{C}_\beta \Delta \hat{\beta}_{ml}^*.
\end{align*}
\]

Using matrix multiplication, we can see that
\[
\begin{align*}
\mathbf{C}_\beta \Delta &= \begin{bmatrix}
0 & (\pi_{11} - \pi_{01}) & \ldots & (\pi_{1k} - \pi_{0k}) \\
0 & (\pi_{21} - \pi_{01}) & \ldots & (\pi_{2k} - \pi_{0k}) \\
\vdots & \vdots & \ddots & \vdots \\
0 & (\pi_{k1} - \pi_{01}) & \ldots & (\pi_{kk} - \pi_{0k})
\end{bmatrix}
\end{align*}
\]

and \( \mathbf{C}_\beta \Delta = \Theta \mathbf{C} \) where
\[
\Theta = \begin{bmatrix}
(\pi_{11} - \pi_{01}) & (\pi_{12} - \pi_{02}) & \ldots & (\pi_{1k} - \pi_{0k}) \\
(\pi_{21} - \pi_{01}) & (\pi_{22} - \pi_{02}) & \ldots & (\pi_{2k} - \pi_{0k}) \\
\vdots & \vdots & \ddots & \vdots \\
(\pi_{k1} - \pi_{01}) & (\pi_{k2} - \pi_{02}) & \ldots & (\pi_{kk} - \pi_{0k})
\end{bmatrix}
\]

so that \( \Theta \) is an invertible matrix. Substituting \( \mathbf{C}_\beta \Delta = \Theta \mathbf{C} \), we get
\[
\begin{align*}
Q_{wc}^0 &= \hat{\beta}_{ml}^* \left( \Theta \mathbf{C}_\beta \right)' \left( \Theta \mathbf{C}_\beta \mathbf{Var}_a(\hat{\beta}_{ml}^*\Theta') \left( \Theta \mathbf{C}_\beta \right)' \right)^{-1} \Theta \mathbf{C}_\beta \hat{\beta}_{ml}^* \\
&= \hat{\beta}_{ml}^* \left( \mathbf{C}_\beta \mathbf{Var}_a(\hat{\beta}_{ml}^*\Theta') \mathbf{C}_\beta \right)^{-1} \Theta \mathbf{C}_\beta \hat{\beta}_{ml}^*.
\end{align*}
\]
\[
\hat{\beta}_{ml}' \mathbf{C}_\beta \mathbf{\Theta} \mathbf{\Theta}^{-1} [ \mathbf{C}_\beta \text{Var}_a(\hat{\beta}^*_{ml}) \mathbf{C}_\beta' ]^{-1} \mathbf{\Theta}^{-1} \mathbf{C}_\beta \hat{\beta}^*_{ml} \\
= \hat{\beta}_{ml}' \mathbf{C}_\beta' [ \mathbf{C}_\beta \text{Var}_a(\hat{\beta}^*_{ml}) \mathbf{C}_\beta' ]^{-1} \mathbf{C}_\beta \hat{\beta}^*_{ml} \\
= Q_{wc}^* 
\]

where \( Q_{wc}^* \) is the statistic which is used when \( \hat{\Pi} = \Pi \) and which tests \( H_0: \beta_1^* = \beta_2^* = \ldots = \beta_k^* = 0. \)

We have shown that the statistic \( Q_{wc} \) above will have the same value whether one ignores misclassification or uses the correct matrix of misclassification probabilities in the fitted model. Since the value of \( \Pi \) was not specified, this proof will hold for any \( \hat{\Pi} \). This shows us that \( Q_{wc} \) is invariant to choices of \( \hat{\Pi} \). Further, since the value of \( Q_{wc} \) is invariant, the p-value associated with the test will also be invariant.

Through this proof, we can also see that testing a hypothesis which involves fewer than \( k \) parameters cannot lead to an invariant \( Q_{wc} \). As an example, let \( H_0: \beta_i = 0 \) be the hypothesis of interest in a situation in which there are more than 2 exposure categories, i.e., \( k > 1 \). Then \( \mathbf{C}_\beta \) will have dimensions \( 1 \times (k+1) \), \( \Delta \) will still have dimensions \( (k+1) \times (k+1) \), so that \( \mathbf{C}_\beta \Delta \) will have dimensions \( 1 \times (k+1) \). The matrix \( \mathbf{\Theta C}_\beta \) will also have dimensions \( 1 \times (k+1) \); it follows that the dimensions of \( \mathbf{\Theta} \) will be \( 1 \times 1 \).

On a closer examination of these matrices, we find
\[
\mathbf{C}_\beta = (0 \ 1 \ 0 \ldots \ 0), \text{ so that } \mathbf{C}_\beta \Delta = (0 \ (\pi_{11} - \pi_{01}) \ldots (\pi_{1k} - \pi_{0k})),
\]
which is the second row of \( \Delta \). Now, define \( \mathbf{\Theta} = (\xi_1) \), so that
\[
\mathbf{\Theta C}_\beta = (0 \ \xi_1 \ 0 \ldots \ 0).
\]
Obviously, \( \mathbf{\Theta C}_\beta \) cannot equal \( \mathbf{C}_\beta \Delta \). Therefore, for \( H_0: \beta_i = 0 \), there
is no matrix $\Theta$ such that $C_\beta \Delta = \Theta C_\beta$.

Suppose, now, that $H_0: \beta_1 = \beta_2 = 0$ is the hypothesis of interest for a situation with $k > 2$. In this case, the dimensions of $\Theta$ are $2 \times 2$.

The matrix $C_\beta$ would be

$$C_\beta = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{bmatrix}$$

so that

$$C_\beta \Delta = \begin{bmatrix} 0 & (\pi_{11} - \pi_{01}) & \cdots & (\pi_{1k} - \pi_{0k}) \\ 0 & (\pi_{21} - \pi_{01}) & \cdots & (\pi_{2k} - \pi_{02}) \end{bmatrix}$$

which are the second and third rows of $\Delta$. Now, define

$$\Theta = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix}.$$  

Then,

$$\Theta C_\beta = \begin{bmatrix} 0 & \xi_{11} & \xi_{12} & 0 & \ldots & 0 \\ 0 & \xi_{21} & \xi_{22} & 0 & \ldots & 0 \end{bmatrix}.$$  

Again, $\Theta C_\beta$ cannot equal $C_\beta \Delta$ and there is no matrix $\Theta$ such that $\Theta C_\beta = C_\beta \Delta$.

From the cases for which $H_0: \beta_1 = 0$ and $H_0: \beta_1 = \beta_2 = 0$ are the hypotheses of interest, we can see a trend in the structure of the matrices $\Theta C_\beta$ and $C_\beta \Delta$. In general, for a hypothesis $H_0: \beta_1 = \beta_2 = \ldots = \beta_j = 0$, $j = 1, 2, \ldots, k$, $\Theta$ will have dimensions $j \times j$. $C_\beta$ will take the form $C_\beta = \begin{pmatrix} 0 & I & 0 \\ j \times 1 & j \times j & j \times (k-j) \end{pmatrix}$, so that $C_\beta \Delta$ will consist of the second through the $(j+1)$-th rows of $\Delta$. $\Theta C_\beta$ will then equal

$$\Theta C_\beta = \begin{pmatrix} 0 & \Theta & 0 \\ j \times 1 & j \times j & j \times (k-j) \end{pmatrix}.$$  

$\Theta C_\beta$ will not equal $C_\beta \Delta$ as long as the right hand matrix of 0's
is present in $\Theta C_\beta$. This is true since the right hand columns of $\Delta$, and therefore of $C_\beta \Delta$, do not consist of 0's. The $0$ matrix is present as long as $j < k$. Thus, only when $j = k$, i.e., when $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$, will there exist a matrix $\Theta$ such that $C_\beta \Delta = \Theta C_\beta$.

The existence of $\Theta$ is necessary in showing the invariance of $Q_{wc}$. We can conclude, then, that the Type I level of significance is retained in the presence of misclassification only when an overall test is performed, i.e., when $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$ is tested. The $Q_{wc}$, and therefore the Type I level of significance, varies with $\Pi$ if a test is performed concerning a proper subset of the $k$ betas. It can also be shown that a test of $H_0: \beta_1 = \beta_2 = \ldots = \beta_k$ is not invariant to choices of $\Pi$.

4.9 Relationship Between $\hat{\beta}^*$ and $\hat{\beta}$

In practice, if a researcher determines that misclassification error of exposure is present in the data, she or he can try to adjust for that through the use of models (1.16), (1.18), or (1.19). However, as we have discussed before, the $\Pi$ matrix will most likely not be known and will have to be estimated. In the earlier sections, we examined the relationship between the estimates of $\beta$ obtained when $\hat{\Pi}^0$ is used in the fitted model (i.e., when misclassification error is ignored) and those obtained when the correct $\Pi$ matrix is used in the fitted model. Here, we will investigate the relationship between the estimates of $\beta$ obtained when $\hat{\Pi}$ is used in the fitted model and those obtained when $\Pi$ is used in the fitted model. The following derivations hold for LS,
WLS, and ML estimation methods. Therefore, we will use the general symbols \( \hat{\beta}^0 \), \( \hat{\beta}^\star \), and \( \hat{\beta} \) to represent the estimators of \( \beta^0 \), \( \beta^\star \), and \( \beta \) for any of the three methods.

To begin, define \( \hat{\Delta} \) as

\[
\hat{\Delta} = \begin{bmatrix}
1 & \hat{\pi}_{01} & \hat{\pi}_{02} & \ldots & \hat{\pi}_{0k} \\
0 & (\hat{\pi}_{11} - \hat{\pi}_{01}) & (\hat{\pi}_{12} - \hat{\pi}_{02}) & \ldots & (\hat{\pi}_{1k} - \hat{\pi}_{0k}) \\
0 & (\hat{\pi}_{21} - \hat{\pi}_{01}) & (\hat{\pi}_{22} - \hat{\pi}_{02}) & \ldots & (\hat{\pi}_{2k} - \hat{\pi}_{0k}) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & (\hat{\pi}_{k1} - \hat{\pi}_{01}) & (\hat{\pi}_{k2} - \hat{\pi}_{02}) & \ldots & (\hat{\pi}_{kk} - \hat{\pi}_{0k}) 
\end{bmatrix}
\]

It follows that \( \hat{\beta}^0 = \hat{\Delta} \hat{\beta}^\star \). We saw before that \( \hat{\beta}^0 = \hat{\Delta} \hat{\beta}^\star \). Therefore, \( \hat{\Delta} \hat{\beta} = \hat{\Delta} \hat{\beta}^\star \), which implies \( \hat{\beta} = \hat{\Delta}^{-1} \hat{\Delta} \hat{\beta}^\star \). Let's examine these matrices in the case of two exposure categories:

\[
\Delta = \begin{bmatrix} 1 & \pi_{01} \\ 0 & (\pi_{11} - \pi_{01}) \end{bmatrix} \quad \hat{\Delta} = \begin{bmatrix} 1 & \hat{\pi}_{01} \\ 0 & (\hat{\pi}_{11} - \hat{\pi}_{01}) \end{bmatrix}
\]

\[
\Rightarrow \hat{\Delta}^{-1} = \begin{bmatrix} 1 & -\hat{\pi}_{01} / (\hat{\pi}_{11} - \hat{\pi}_{01}) \\ 0 & 1 / (\hat{\pi}_{11} - \hat{\pi}_{01}) \end{bmatrix}
\]

\[
\Rightarrow \hat{\beta} = \hat{\Delta}^{-1} \hat{\Delta} \hat{\beta}^\star = \begin{bmatrix} \hat{\beta}_{\alpha}^\star + \frac{\hat{\pi}_{11} \hat{\pi}_{01} - \hat{\pi}_{01} \hat{\pi}_{11}}{\hat{\pi}_{11} - \hat{\pi}_{01}} \hat{\beta}_1^\star \\ \frac{\pi_{11} - \pi_{01}}{\hat{\pi}_{11} - \hat{\pi}_{01}} \hat{\beta}_1^\star \end{bmatrix}
\]

(4.15)
This is the same relationship found to hold between the parameters $\beta = \mathbb{E}(\hat{\beta}_1)$ and $\beta^*$ in Section 3.2.2. We would expect the two relationships to be identical. Expression (4.15) was derived based on the equality $\hat{\beta}^0 = \Delta \hat{\beta}^*$. This same equality was found to hold for the parameters. Therefore, having the relationship of expression (4.15) hold for $\beta$ and $\beta^*$, as well, is consistent with the theory.

Let us further investigate the expression $\Delta^{-1} \Delta$. Recall that $\Pi_1 = \hat{\Pi}^{-1}_1 \Delta$, so that $\Delta = \hat{\Pi}^{-1}_1 \Pi_1$. We can also show that $\hat{\Pi}_1 = \hat{\Pi}^{-1}_1 \Delta$, so that $\Delta = \hat{\Pi}^{-1}_1 \hat{\Pi}_1$ and $\hat{\Delta} = \hat{\Pi}^{-1}_1 \hat{\Pi}_1$. It follows that $\Delta^{-1} \Delta = \hat{\Pi}_1^{-1} \Pi_1$.

Therefore, for each type of estimation method considered, $\hat{\beta} = \hat{\Pi}_1^{-1} \Pi_1 \beta^*$. Further, conditioning on $\hat{\Pi}_1$, the expected value of $\hat{\beta}$ is $\Pi_1^{-1} \Pi_1 \beta^*$ (asymptotically, in the case of WLS and ML estimation). This was shown in Chapter 3 for LS and WLS estimators. The proof given here applies to all three types of estimation methods.

Returning to the case of $k=1$, by examining the expression

$$
\hat{\beta}_1 = \hat{\beta}_1^* \left[ \frac{\pi_{11} - \pi_{01}}{\pi_{11} - \pi_{01}} \right]
$$

we can see that it is the disparity between the difference $(\pi_{11} - \pi_{01})$ and its estimate, not the disparity between the values of $\pi_{11}$ and $\pi_{01}$ and their estimates, per se, which determines the extent of the difference between the two estimates of $\beta$. For example, if

$$
\Pi_1 = \begin{bmatrix} 1 & 1/4 \\ 1 & 3/4 \end{bmatrix} \quad \text{and} \quad \hat{\Pi}_1 = \begin{bmatrix} 1 & 1/6 \\ 1 & 2/3 \end{bmatrix},
$$

then $\hat{\beta}_1 = \hat{\beta}_1^*$.
even though $\hat{\pi}_{11} \neq \hat{\pi}_{11}$ and $\hat{\pi}_{01} \neq \hat{\pi}_{01}$, since

$$(3/4 - 1/4) = (2/3 - 1/6) = 1/2.$$ However, if $\Pi_1 = \begin{bmatrix} 1 & 1/3 \\ 1 & 2/3 \end{bmatrix}$, then $\hat{\beta}_1 = 3/2 \hat{\beta}_1^\star$.

Since $\pi_{01} = (1 - \pi_{00})$, the condition $\pi_{11} - \pi_{01} = \hat{\pi}_{11} - \hat{\pi}_{01}$ holds when $\pi_{11} + \pi_{00} = \hat{\pi}_{11} + \hat{\pi}_{00}$ holds. $\pi_{11}$ and $\pi_{00}$ can be thought of the probabilities of correctly classifying subjects. Therefore, with two exposure categories, $\hat{\beta}_1 = \hat{\beta}_1^\star$ as long as the average of the estimated and true probabilities of correctly classifying subjects are equal.

Considering a case of three exposure categories, it can be shown that $\hat{\beta}_1 = \hat{\beta}_1^\star$ and $\hat{\beta}_2 = \hat{\beta}_2^\star$ when all four of the following conditions hold:

\[
\begin{align*}
\pi_{11} - \pi_{01} &= \hat{\pi}_{11} - \hat{\pi}_{01} \\
\pi_{21} - \pi_{01} &= \hat{\pi}_{21} - \hat{\pi}_{01} \\
\pi_{22} - \pi_{02} &= \hat{\pi}_{22} - \hat{\pi}_{02} \\
\pi_{12} - \pi_{02} &= \hat{\pi}_{12} - \hat{\pi}_{02}.
\end{align*}
\]

These conditions are less easily interpreted than the condition for the two exposure category case. They perhaps become clearer upon returning to the structure of the matrix $\Pi_1$, shown below:

$$\Pi_1 = \begin{bmatrix} 1 & \pi_{01} & \pi_{02} \\
1 & \pi_{11} & \pi_{12} \\
1 & \pi_{21} & \pi_{22} \end{bmatrix}$$

The four equalities given above are comparisons of the misclassification probabilities between the second and lowest and between the third and lowest categories.
Again, we see that it is not the values of the $\pi_{ij}$'s themselves which affect the degree to which the $\hat{\beta}$'s differ, but the differences between certain $\pi_{ij}$'s. As an example, consider a classification scheme with three exposure categories. The misclassification probabilities are given in $\Pi_1$; an estimated set is given in $\hat{\Pi}_1$:

$$\Pi_1 = \begin{bmatrix} 1 & .0420 & .0336 \\ 1 & .5814 & .1860 \\ 1 & .0610 & .8293 \end{bmatrix} \quad \hat{\Pi}_1 = \begin{bmatrix} 1 & .05 & .05 \\ 1 & .5894 & .2024 \\ 1 & .069 & .8457 \end{bmatrix}$$

When these two matrices are used in regression analysis, the resulting estimates for $\beta_1$ and $\beta_2$ are the same. This occurs since the four conditions hold:

$$0.5814 - 0.0420 = 0.5394 = 0.5894 - 0.05$$
$$0.0610 - 0.0420 = 0.019 = 0.069 - 0.05$$
$$0.8293 - 0.0336 = 0.7957 = 0.8457 - 0.05$$
$$0.1860 - 0.0336 = 0.1524 = 0.2024 - 0.05$$

In the next chapter, we will examine how the estimates of $\beta$ are affected when varying degrees of misclassification are assumed. This is done through the use of examples in an effort to see how sensitive results of an analysis can be to misclassification of the data.

4.10 Removal of Restriction on Data for LS Estimation

This final section of Chapter 4 deals exclusively with LS estimation. The theory developed in this and the previous chapters for
LS estimators was based on the assumption of equal numbers of subjects in each classified exposure category. This assumption was necessary to express \( \Pi \) as
\[
\Pi = \begin{bmatrix}
\Pi_1 \\
\Pi_2 \\
\vdots \\
\Pi_k
\end{bmatrix}.
\]

Such an assumption, however, places an unrealistic restriction on the data. Here, we will investigate the situation in which the assumption is violated. Specifically, we will determine whether the results for LS estimation derived in Chapter 3 and earlier sections of this chapter are still valid.

Suppose that the number of subjects in the \( i \)-th classified exposure category is \( S_i \), \( i=0,1,\ldots,k \). Let us now define the total number of subjects as \( S' = \sum_{i=0}^{k} S_i \). If the assumption holds, then \( S_i = S \) for all \( i=0,1,\ldots,k \), and \( S' = (k+1)S \). Here we are assuming, however, that the \( S_i \)'s are not all equal.

If the response variables are ordered such that
\[
Y=(Y_{01},Y_{02},\ldots,Y_{0S_0},Y_{11},Y_{12},\ldots,Y_{1S_1},\ldots,Y_{k1},Y_{k2},\ldots,Y_{kS_k})',
\]
then the \( \Pi \) matrix will have the form
Expression (4.5) implies that $\Pi = \hat{\Pi}^0 \Delta$ for the case in which $S_i = S$, $i = 0, 1, \ldots, k$. This relationship also holds if $S_i \neq S$ for all $i = 0, 1, \ldots, k$. Further, the relationship $\hat{\Pi} = \hat{\Pi}^0 \Delta$ holds in this case as well. We will use this second relationship to show the invariance of $\hat{R}$, and therefore of $\hat{\gamma}_{1s}$, as follows.

First, consider the matrix expression

$$
\hat{\Pi} (\hat{\Pi}^0 \hat{\Pi})^{-1} \hat{\Pi} = \hat{\Pi}^0 \Delta \left[ (\hat{\Pi}^0 \Delta) \left( (\hat{\Pi}^0 \Delta)^{-1} (\hat{\Pi}^0 \Delta) \right)^{-1} \right]
$$

which is invariant to choices of $\hat{\Pi}$. Since $\hat{R} = [I - \hat{\Pi} (\hat{\Pi}^0 \hat{\Pi})^{-1} \hat{\Pi}]$, we can see that $\hat{R}$ is also invariant to choices of $\hat{\Pi}$. The invariance of $\hat{R}$ leads directly to the invariance of $\hat{\gamma}_{1s}$ and of $\text{Var}(\hat{\gamma}_{1s})$. 

124
To prove the invariance of the predicted values, we need to show that \( \hat{\Pi}^t_1 \) is invariant to changing degrees of misclassification. This is easily done using the derivation presented above.

\[
\hat{\Pi}^t_1 = \hat{\Pi} (\hat{\Pi}' \hat{\Pi})^{-1} \hat{\Pi}' (Y - V \hat{Y})_1
\]

\[
= \hat{\Pi} (\hat{\Pi}' \hat{\Pi}^0)^{-1} \hat{\Pi}' (Y - V \hat{Y})_1
\]

It follows that the vector of predicted values, \( \hat{Y} = \hat{\Pi}^t_1 + V \hat{Y} \), is also invariant. The invariance of \( \hat{Y} \) is used to conclude that \( \text{Var}(\hat{Y}_1) \), \( R^2 \), and the F statistic are all invariant to choices of \( \hat{\Pi} \). In fact, all of the results found for LS estimation in Chapter 3 can be shown to hold in this scenario.

The conclusions reached in Sections 4.4, 4.5, 4.6, 4.7, and 4.8 concerning LS estimation hinged on the relationship found in expression (4.9). If we can show that this relationship holds when \( S_1 \) for \( i=0,1,...,k \), then we can conclude that the results in this chapter for LS estimators hold in this case as well.

As we saw in Section 4.4,

\[
\hat{\beta}_1 = (\Pi' \Pi)^{-1} \Pi' (Y - V \hat{Y})_1
\]

Using the expression \( \Pi = \hat{\Pi} \Delta \), this becomes

\[
\hat{\beta}_1 = \Delta^{-1} (\hat{\Pi}' \hat{\Pi}^0)^{-1} \hat{\Pi}' (Y - V \hat{Y})_1
\]

Since \( \hat{Y}_1 \) is invariant to choices of \( \hat{\Pi} \), \( \hat{Y}_1 = \hat{Y}_{10} \), so that

\[
\hat{\beta}_1 = \Delta^{-1} \hat{\Pi} \hat{\beta}_1 = \Delta^{-1} \hat{\beta}_1
\]

which is the relationship found in expression (4.9). This shows us that the results found in the earlier sections of this chapter do not
require that there be equal numbers of subjects in the classified exposure categories. Since this was also found to be true of the results in Chapter 3, we will dispense with this unnecessary requirement.
CHAPTER 5
NUMERICAL APPLICATIONS

5.1 Introduction

Numerical examples are used in this chapter to investigate two issues regarding the use of the models developed in Chapter 1. First, the question of when models (1.16) and (1.18) are appropriate is addressed. The theoretical implications of this were examined in Section 1.11. The examples examined here point to a number of conditions which can help ensure that the models will perform reasonably well.

In the second part of this chapter, models (1.16), (1.18), and (1.19) are each used to analyze a different data set. Each model is fit assuming several different sets of misclassification probabilities. Through these analyses, the sensitivity of the results to varying assumed degrees of misclassification is probed.

5.2 Conditions for Appropriateness of Models

The theory developed in earlier chapters assumes that models (1.16) and (1.18) are appropriate to a given set of risks or rates and
misclassification probabilities. Since this is not always the case, it is valuable to know under what conditions one can expect these models to perform well.

As has been mentioned, model (1.16) assumes that model (1.6) holds within each stratum. This implies that expression (1.21) holds within each stratum. Likewise, model (1.18) assumes that model (1.7) and, therefore, expression (1.22) hold within each stratum. By determining the conditions needed for models (1.6) and (1.7) to hold within a given stratum, we will arrive at the conditions needed for models (1.16) and (1.18) to hold. Therefore, for the time being, let us restrict our enquiry to the fit of models (1.6) and (1.7) in a single stratum.

Consider a situation with no misclassification of exposure and only one stratum. In this case, expressions (1.21) and (1.22) hold perfectly, as do models (1.6) and (1.7). Here, the misclassification probabilities are such that \( \pi_{ij} = 1 \) for all \( i=j \) and \( \pi_{ij} = 0 \) otherwise, for \( i,j=0,1,\ldots,k \) (i.e., \( \Pi = \Pi^0 \)). Intuitively, it seems that for a given set of true risks, the closer the pattern of values in the matrix \( \Pi \) are to the pattern of values in \( \Pi^0 \), the better the fit of models (1.6) and (1.7).

5.2.1 Logistic Regression.

Let us consider, first, the case of logistic regression. The examples which follow indicate that, in general, the above statement is true with regard to the fit of model (1.6). When all of the \( \pi_{ij} \)'s with \( i=j \) are equal (i.e., \( \pi_{00}=\pi_{11}=\ldots=\pi_{kk} \)), as their common value decreases from 1, the quality of the estimates resulting from the use
of model (1.6) decreases. The "quality" of an estimate, in this context, refers to how close that estimate is to the estimate from the (hypothetical) true data. It cannot be said that a lower degree of misclassification will always lead to higher quality estimates; however, in general, this will be the case. Also, for a given set of misclassification probabilities, the quality of estimates diminishes as the OR* j 's increase.

These points are illustrated in the following tables. The values in Table 5.1 were calculated for a given OR* 1 with k=1 and S=1. They illustrate the effect that varying sets of misclassification probabilities have on the quality of the estimate produced assuming model (1.6) and, therefore, model (1.16). The value of OR* 1 is set at 2.111 so that (ln OR* 1 )=β 1 *=.7472. The estimator from model (1.16) is (ln OR* 1 )=β 1 * which is the estimator derived when k=1. We are considering these π i j 's as known and the data as population data. Therefore, we would expect β 1 *=β 1 *. The quality of the estimate is determined by the absolute differences |β 1 *-β 1 *| and |OR* 1 -OR* 1 |. The third column contains (π 11 +π 00 ) which is used as a measure of the total degree of non-misclassification involved. The higher the value in this column, the lower the degree of misclassification; of course, (π 11 +π 00 )=2 corresponds to no misclassification.
Table 5.1

Quality of estimates for various $\pi_{ij}$'s
given $OR_1^* = 2.111$ ($\beta_1^* = .7472$)

| $\pi_{00}$ | $\pi_{11}$ | $(\pi_{00}+\pi_{11})$ | $|\hat{OR}_1^*-OR_1^*|$ | $|\hat{\beta}_1^*-\beta_1^*|$ |
|-----------|-----------|------------------------|---------------------------|---------------------------|
| .9500     | .9500     | 1.9000                 | .0112                     | .0053                     |
| .9474     | .9048     | 1.8522                 | .0057                     | .0027                     |
| .9048     | .9474     | 1.8522                 | .0372                     | .0178                     |
| .9000     | .9000     | 1.8000                 | .0210                     | .0100                     |
| .8889     | .8182     | 1.7071                 | .0144                     | .0032                     |
| .8182     | .8889     | 1.7071                 | .0621                     | .0299                     |
| .8500     | .8500     | 1.7000                 | .0308                     | .0147                     |
| .8421     | .8095     | 1.6816                 | .0168                     | .0080                     |
| .8000     | .8000     | 1.6000                 | .0364                     | .0174                     |
| .8571     | .6923     | 1.5494                 | .0452                     | .0212                     |
| .7600     | .7600     | 1.5200                 | .0412                     | .0197                     |
| .7500     | .6667     | 1.4167                 | .0061                     | .0029                     |
| .7000     | .7000     | 1.4000                 | .0470                     | .0225                     |
| .6500     | .6500     | 1.3000                 | .0500                     | .0239                     |
| .6000     | .6000     | 1.2000                 | .0532                     | .0255                     |

Evaluation of the results in this table indicates three characteristics of the relationship between the values of the $\pi_{ij}$'s and the quality of the estimates. First, when $\pi_{00}=\pi_{11}$, as their common value decreases, the quality of the estimate decreases. By comparing the table entries corresponding to such sets of $\pi_{ij}$'s, we can see that as $\pi_{00}=\pi_{11}$ takes on values from .95 to .60, the differences $|\hat{\beta}_1^*-\beta_1^*|$ and $|\hat{OR}_1^*-OR_1^*|$ increase steadily.

Second, in general, the higher the values of $\pi_{00}$ and $\pi_{11}$, the closer the estimate of $OR_1^*$ is to the true value. Eight of the nine closest
estimates correspond to misclassification probabilities with both \( \pi_{00} \geq 0.80 \) and \( \pi_{11} \geq 0.80 \).

Third, the quality of the estimate will probably be better if \( \pi_{00} > \pi_{11} \), than if \( \pi_{00} = \pi_{11} \) or if \( \pi_{00} < \pi_{11} \). First, compare the first two sets of \( \pi_{ij} \)'s in the table. Although the first set has a lower total degree of misclassification, the differences in columns four and five are greater in magnitude than the differences corresponding to the second set of \( \pi_{ij} \)'s. The same is true for the fourth and fifth sets of misclassification probabilities.

Second, compare the second and third sets of misclassification probabilities. Notice that the values of the \( \pi_{ij} \)'s are switched between the two sets. The value of \( \pi_{11} \) in the first is equal to the value of \( \pi_{00} \) in the second, and vice versa. Obviously, the total degree of misclassification is equal for the two sets. The qualities of the resulting estimates, however, are quite different. The second set has the smallest differences found in the table. The differences corresponding to the third set are over four times those corresponding to the second set, but are still quite small in size.

A similar phenomenon is found when the fifth and sixth sets of \( \pi_{ij} \)'s are compared. The values of \( \pi_{00} \) and \( \pi_{11} \) are also switched between these sets. The differences corresponding to the sixth set are not only much larger than those corresponding to the fifth set, they are the largest in the table.

Unlike the cases in which \( \pi_{00} = \pi_{11} \), when the two misclassification probabilities are not equal, there is no definite downward trend in the quality of the estimate as the values of the \( \pi_{ij} \)'s decrease. This is
poignantly illustrated by the set of probabilities $\pi_{00}=.7500$ and $\pi_{11}=.6667$. These probabilities correspond to the second smallest difference $|\hat{\beta}^{*}-\beta^{*}|$ of all entries in the table.

As a final, but important point, let us note that the differences in columns four and five are all quite small. The biggest difference leads to a value of $\hat{OR}_{1}^{*}=2.0489$ which is not very far from the true value of 2.111. In general, we can say that the model performs reasonably well for all sets of $\pi_{ij}$'s in Table 5.1, although it does perform better for some than for others.

Table 5.2 is used to illustrate the point that, all else being equal, the larger the true odds ratio, the lower the quality of the estimates. Rather than varying the misclassification probabilities for a given $OR_{1}^{*}$, as we did for Table 5.1, here we vary the $OR_{1}^{*}$'s for a given set of misclassification probabilities. Again, $k=1$ and $S=1$. Each of the estimates in Table 5.2 was calculated using $\pi_{00}=\pi_{11}=.80$. As the values of the $OR_{1}^{*}$'s increase, the absolute differences $|\hat{OR}_{1}^{*}-OR_{1}^{*}|$ and $|\hat{\beta}^{*}-\beta^{*}|$ also increase, as does the difference $|\hat{OR}_{1}^{*}-OR_{1}^{*}|$ relative to the value of $OR_{1}^{*}$.

**Table 5.2**

| $OR_{1}^{*}$ | $|\hat{\beta}^{*}-\beta^{*}|$ | $|\hat{OR}_{1}^{*}-OR_{1}^{*}|$ | $\frac{|\hat{OR}_{1}^{*}-OR_{1}^{*}|}{x100}$ |
|-----------|----------------|----------------|-------------------|
| 1.65      | .0055          | .0004          | .253              |
| 2.11      | .0174          | .036           | 1.721             |
| 2.59      | .0346          | .088           | 3.396             |
| 4.75      | .1301          | .580           | 12.212            |

132
Table 5.3 includes several examples with k=2 and S=1 which are used to investigate whether the trends found for k=1 carry over to this situation. Table 5.3 was constructed in the same fashion as Table 5.1. In this case, OR₁* and OR₂* are given (OR₁*=2.111 and OR₂*=4.75). The misclassification probabilities are then varied and the resulting estimates of OR₁* and OR₂* are compared with the true parameters. In addition, the estimates produced from model (1.16) (β₁* and β₂*) are compared with the true values (β₁*=.7472 and β₂*=1.5581). For each example, the πᵢⱼ’s are presented in a table of the form:

<table>
<thead>
<tr>
<th>i=0</th>
<th>j=0</th>
<th>j=1</th>
<th>j=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>π₀₀</td>
<td>π₀₁</td>
<td>π₀₂</td>
<td></td>
</tr>
<tr>
<td>i=1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>π₁₀</td>
<td>π₁₁</td>
<td>π₁₂</td>
<td></td>
</tr>
<tr>
<td>i=2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>π₂₀</td>
<td>π₂₁</td>
<td>π₂₂</td>
<td></td>
</tr>
</tbody>
</table>

The results shown in this table indicate that the trends found when k=1 also hold when k=2. In general, as the probabilities on the main diagonal increase, the quality of the estimates increases. Also, by comparing the third and fourth sets of πᵢⱼ’s, we can see that the quality of the estimates is probably better when π₀₀+π₁₁+π₂₂. Since each of the diagonal elements in the third set is larger than each of the diagonal elements in the fourth set, one would expect closer estimates from the third set. This, however, is not the case. Therefore, we conclude that the larger difference is due to the equality of the πᵢⱼ’s on the main diagonal. Also, the differences in
this table are small enough to indicate that the model works well for all sets of $\pi_{ij}$'s considered.

**Table 5.3**

$|\hat{OR}_{j*}-OR_{j*}|$ (and $|\hat{\beta}_{j*}-\beta_{j*}|$) $j=1,2$ for various $\pi_{ij}$'s given $OR_{1*}=2.111$ and $OR_{2*}=4.750$ ($\beta_{1*}=.7472$ and $\beta_{2*}=1.5581$)

<table>
<thead>
<tr>
<th>$\pi_{ij}$'s</th>
<th>Absolute Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$j=1$</td>
</tr>
<tr>
<td>.9375</td>
<td>.0521</td>
</tr>
<tr>
<td>.0600</td>
<td>.9000</td>
</tr>
<tr>
<td>.0385</td>
<td>.0481</td>
</tr>
<tr>
<td>.9677</td>
<td>.0215</td>
</tr>
<tr>
<td>.0714</td>
<td>.8889</td>
</tr>
<tr>
<td>.0202</td>
<td>.0202</td>
</tr>
<tr>
<td>.9000</td>
<td>.0600</td>
</tr>
<tr>
<td>.0500</td>
<td>.9000</td>
</tr>
<tr>
<td>.0300</td>
<td>.0700</td>
</tr>
<tr>
<td>.8649</td>
<td>.1081</td>
</tr>
<tr>
<td>.1237</td>
<td>.8247</td>
</tr>
<tr>
<td>.0724</td>
<td>.0905</td>
</tr>
</tbody>
</table>

134
Finally, Table 5.4 investigates the situation for k=3. These examples, being so few, cannot be used to say anything definitive regarding trends in the relationship between the quality of the estimates and the choice of misclassification probabilities. Instead, these examples are used to support the findings of Tables 5.1 and 5.3, and to give an idea of how close the estimates will be to the true values in these particular examples. The true odds ratios are set at OR$_1^* = 2.111$, OR$_2^* = 4.75$, and OR$_3^* = 6.333$ ($\beta_1^* = .7472$, $\beta_2^* = 1.5581$, and $\beta_3^* = 1.8458$). The $\pi_{ij}$'s are presented in a table of the form
\[
\begin{array}{cccc}
  & j=0 & j=1 & j=2 & j=3 \\
 i=0 & \pi_{00} & \pi_{01} & \pi_{02} & \pi_{03} \\
 i=1 & \pi_{10} & \pi_{11} & \pi_{12} & \pi_{13} \\
 i=2 & \pi_{20} & \pi_{21} & \pi_{22} & \pi_{23} \\
 i=3 & \pi_{30} & \pi_{31} & \pi_{32} & \pi_{33} \\
\end{array}
\]

The same trends seen in Tables 5.1 and 5.3 can be seen in Table 5.4. Firstly, as the degree of misclassification decreases, the quality of the estimates increase. Secondly, none of the sets of \(\pi_{ij}\)'s has equal probabilities on the main diagonal and, in turn, none of the differences \(|\hat{\beta}_j - \hat{\beta}_j|\) for the first two sets is greater than .1. This is true even though two of the main diagonal probabilities in the second set are less than .8. The results in this table, therefore, support the hypothesis that unequal, relatively high probabilities of correct classification will help ensure high quality estimates.

Also, we can compare across these three tables to examine the effect that a greater number of exposure categories has on the quality of the estimates. For instance, in Table 5.1 when \(\pi_{00}=\pi_{11}=.80\), \(|\hat{\text{OR}}_{1^*}-\hat{\text{OR}}_{1^*}|=.036\). When \(k=2\) and \(\pi_{00}=\pi_{11}=.80\), however, \(|\hat{\text{OR}}_{1^*}-\hat{\text{OR}}_{1^*}|=.226\) even though the value of \(\text{OR}_{1^*}\) is the same. The differences \(|\hat{\text{OR}}_{1^*}-\hat{\text{OR}}_{1^*}|\) in Table 5.1 never exceed .06, even with high degrees of misclassification. These differences in Tables 5.3 and 5.4, however, exceed that number when there are moderate to low degrees of misclassification. One would contend that the higher the number of \(\beta^*\)'s being estimated, the lower the quality of the estimates. These tables support this contention. It is
Table 5.4

\[ |\hat{\beta}_j^* - \beta_j^*| \ (\text{and } |\hat{\beta}_j^* - \beta_j^*|) \ j=1,2,3 \text{ for various } \pi_{ij}\text{'s}

given OR_{1^*}=2.111, OR_{2^*}=4.750, \text{ and } OR_{3^*}=6.333

(\hat{\beta}_{1^*}=7.472, \hat{\beta}_{2^*}=1.5581, \text{ and } \hat{\beta}_{3^*}=1.8458)

<table>
<thead>
<tr>
<th>\pi_{ij}\text{'s}</th>
<th>Absolute Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>j=1</td>
</tr>
<tr>
<td>.9422</td>
<td>.9428</td>
</tr>
<tr>
<td>.0607</td>
<td>.9109</td>
</tr>
<tr>
<td>.0384</td>
<td>.0384</td>
</tr>
<tr>
<td>.0193</td>
<td>.0193</td>
</tr>
<tr>
<td>.8511</td>
<td>.0851</td>
</tr>
<tr>
<td>.1132</td>
<td>.7547</td>
</tr>
<tr>
<td>.0566</td>
<td>.0755</td>
</tr>
<tr>
<td>.0213</td>
<td>.0426</td>
</tr>
<tr>
<td>.7609</td>
<td>.1304</td>
</tr>
<tr>
<td>.1667</td>
<td>.6481</td>
</tr>
<tr>
<td>.0741</td>
<td>.1111</td>
</tr>
<tr>
<td>.0435</td>
<td>.0652</td>
</tr>
</tbody>
</table>

Important to note, also, that even with k=3 the differences are relatively small, especially for low degrees of misclassification. This leads us to believe that a good quality estimate can be obtained with a large number of exposure categories as long as the degree of misclassification is not too high and unequal between exposure.

137
categories.

Therefore, we conclude that model (1.6) will hold reasonably well for a given stratum when three conditions hold. The first is that the degrees of misclassification are low so that the smallest \( \pi_{ij} \) with \( i=j \) is greater than or equal to .80. The second condition calls for low magnitudes of the true odds ratios. Exactly how low these magnitudes should be depends on the misclassification probabilities involved. The higher the degree of non-misclassification, the higher the magnitudes can be to obtain quality estimates. In general, an odds ratio below three will be estimated well. Finally, the probabilities of correct classification should not be equal for all exposure categories.

The first condition is necessary in guaranteeing an estimate reasonably close to the true estimate. If the the second condition is violated, the estimates produced may still be a good approximation of the true estimate as long as the first and third conditions are not violated. For example, we have seen that for probabilities of correct classification which are unequal and .80 or above, high quality estimates can be obtained of odds ratios with values of six or below.

The third condition is an added guarantee of quality estimates. It becomes especially important as the degree of misclassification increases.

5.2.2 Poisson Regression

We can see the same conditions hold in the case of Poisson regression. First, Table 5.5 is analogous to Table 5.1. Here, the
value of IDR* is given for a situation with k=1 and S=1
(IDR1*=2.25). Various sets of misclassification probabilities are
assumed; estimates were calculated for each set using model (1.7).
The absolute differences between the true parameters and their
estimates are shown in the table.

As we saw in the case of logistic regression, the absolute
differences between the true values of the parameters and their
estimates are all small. Further, these differences tend to decrease
as the degree of misclassification decreases.

| Table 5.5 |
| Quality of estimates for various $\pi_{ij}$'s given IDR1*=2.25 ($\beta_1*=.8109$) |
|-----------|-----------------|------------|-------------|------------------|
| $\pi_{00}$ | $\pi_{11}$     | $(\pi_{00}+\pi_{11})$ | $|\hat{\beta}_1*-\beta_1*|$ | $|\text{IDR}_1*-\text{IDR}_1*|$ |
| .9794      | .9515          | 1.9309     | .0035       | .0078            |
| .9515      | .9794          | 1.9309     | .0147       | .0329            |
| .9000      | .9000          | 1.8000     | .0159       | .0356            |
| .8947      | .8571          | 1.7518     | .0067       | .0151            |
| .8571      | .8947          | 1.7518     | .0300       | .0686            |
| .8000      | .8000          | 1.6000     | .0276       | .0613            |
| .7895      | .7619          | 1.5514     | .0212       | .0473            |

Another similarity to the logistic case is seen through a
comparison of the first two sets of misclassification probabilities.
The values of $\pi_{00}$ and $\pi_{11}$ are switched between these two sets. The
differences corresponding to the first set are much smaller than those
corresponding to the second set. The same holds for the fourth and
fifth sets. Compare, also, the fifth and sixth sets. Notice that,
although the total degree of misclassification is higher, the quality of
estimates for the fifth set is poorer than that for the sixth set. These comparisons indicate that the quality of the estimates is greatest when $\pi_{00} > \pi_{11}$. The quality decreases when $\pi_{00} = \pi_{11}$, and decreases further when $\pi_{11} > \pi_{00}$.

Table 5.6 for Poisson regression is analogous to Table 5.2 for logistic regression. Again, the misclassification probabilities are set at $\pi_{00} = \pi_{11} = .8$; $k=1$, and $S=1$. The value of IDR$_1^{**}$ is varied and the quality of the estimates is examined. As in Table 5.2, these results indicate that the quality decreases as the value of IDR$_1^{**}$ increases, both in terms of the absolute differences and in terms of the differences relative to the value of IDR$_1^{**}$.

**Table 5.6**

Quality of estimates for various IDR$_1^{**}$'s given $\pi_{00} = \pi_{11} = .80$

| IDR$_1^{**}$ | $|\hat{\beta}_1^{**} - \beta_1^{**}|$ | $|\hat{\text{IDR}}_1^{**} - \text{IDR}_1^{**}|$ | IDR$_1^{**}$ x 100 |
|-------------|---------------------------------|---------------------------------|-----------------|
| 1.50        | .004                            | .005                            | .333            |
| 2.25        | .028                            | .061                            | 2.711           |
| 3.00        | .025                            | .073                            | 2.433           |
| 4.50        | .164                            | .697                            | 15.089          |
| 9.00        | .454                            | 3.284                           | 36.489          |

Tables 5.7 and 5.8 examine the cases of $k=2$ and $k=3$, respectively. For both tables, the values of the IDR$_j^{**}$'s were given and the misclassification probabilities varied. Again, the qualities of the estimates from the two tables are quite good, with the quality generally increasing as the degree of misclassification decreases.
Table 5.7

$|\hat{f}^{*}_{j} - I^*_j| \ (\text{and} \ |\hat{g}^{*}_{j} - g^{*}_{j}|), \ j=1,2, \ for \ various \ sets \ of \ \pi_{ij}'s$

given $I^*_1=2.25$ and $I^*_2=3.00 \ (g^{*}_{1}=.8109 \ and \ g^{*}_{2}=.10986)$

<table>
<thead>
<tr>
<th>$\pi_{ij}$'s</th>
<th>$j=1$</th>
<th>$j=2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>.0667</td>
<td>.8889</td>
<td>.0444</td>
</tr>
<tr>
<td>.0256</td>
<td>0.000</td>
<td>.9744</td>
</tr>
<tr>
<td>.9524</td>
<td>.0226</td>
<td>.0251</td>
</tr>
<tr>
<td>.0353</td>
<td>.8941</td>
<td>.0706</td>
</tr>
<tr>
<td>.0134</td>
<td>.0241</td>
<td>.9626</td>
</tr>
<tr>
<td>.8780</td>
<td>.0976</td>
<td>.0244</td>
</tr>
<tr>
<td>.0789</td>
<td>.8421</td>
<td>.0789</td>
</tr>
<tr>
<td>.0244</td>
<td>.0976</td>
<td>.8780</td>
</tr>
<tr>
<td>.8718</td>
<td>.1026</td>
<td>.0256</td>
</tr>
<tr>
<td>.1000</td>
<td>.8000</td>
<td>.1000</td>
</tr>
<tr>
<td>.0488</td>
<td>.0976</td>
<td>.8537</td>
</tr>
</tbody>
</table>
Let us conclude that both models (1.6) and (1.7) perform reasonable well in the examples we have seen here. The quality of the estimates resulting from the use of these models will generally improve when the degree of misclassification is low, when the probabilities of correct classification vary (at least a little) from category to category, and when a small number of $\beta_j$'s are being estimated. Further, since models (1.16) and (1.18) assume that models (1.6) and (1.7) hold within each stratum, respectively, these conditions relate to the appropriateness of models (1.16) and (1.18) when any number of strata are involved.
Table 5.8

\[ |\hat{\text{IDR}}_j - \text{IDR}_j^*| \ (\text{and} \ |\hat{\beta}_j - \beta_j^*|), \ j=1, 2, 3, \ \text{for various} \ \pi_{ij}'s \]

given \text{IDR}_1^* = 2.25, \ \text{IDR}_2^* = 3.00, \ \text{and} \ \text{IDR}_3^* = 5.00

(\beta_1^* = .8109, \ \beta_2^* = 1.0986, \ \text{and} \ \beta_3^* = 1.6094)

<table>
<thead>
<tr>
<th>\pi_{ij}'s</th>
<th>j=1</th>
<th>j=2</th>
<th>j=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>.9375</td>
<td>.0260</td>
<td>.0260</td>
<td>.0104</td>
</tr>
<tr>
<td>.0505</td>
<td>.9091</td>
<td>.0253</td>
<td>.0152</td>
</tr>
<tr>
<td>.0238</td>
<td>.0476</td>
<td>.8571</td>
<td>.0714</td>
</tr>
<tr>
<td>.0250</td>
<td>.0250</td>
<td>.0500</td>
<td>.9000</td>
</tr>
</tbody>
</table>

| .8247     | .1031| .0515| .0206|
| .0995     | .7960| .0746| .0299|
| .0476     | .0476| .7619| .1429|
| .0513     | .0513| .0769| .8205|

| .7568     | .1351| .0541| .0541|
| .1395     | .6512| .1163| .0930|
| .0930     | .1163| .6512| .1395|
| .0541     | .0541| .0351| .7568|

5.3 Numerical Examples

There are two different scenarios which may occur when there is a possibility of misclassification of exposure in collected data. In the
first, the researcher knows the data is misclassified. In some way, she or he estimates the \( \pi_{ij} \)’s and then uses these estimates to calculate \( \hat{\beta} \), which will be a biased estimator of the true parameter \( \beta^{*} \), unless the researcher estimates the \( \pi_{ij} \)’s perfectly (an extremely unlikely event). In our numerical examples up to this point, we have assumed that the \( \pi_{ij} \)’s were known.

In the second scenario, the researcher knows or suspects there is misclassification, but does not (or perhaps cannot) estimate \( \Pi \). Instead, she or he first calculates the estimate \( \hat{\beta} \) of \( \beta^{*} \) assuming there is no misclassification (i.e., using \( \hat{\Pi} = \hat{\Pi}^{0} \)). Then the researcher asks the question: How much would this estimate change if \( \hat{\Pi} \) had some other structure? The researcher addresses this question by assuming several different structures for \( \hat{\Pi} \) and then comparing the resulting \( \hat{\beta} \)’s. This will give an indication of how sensitive the results are to varying assumed degrees of misclassification.

Under this scenario, the model being fit is: \( \mathbf{E}(\mathbf{Y}) = \hat{\Pi} \beta + \mathbf{V}_Y \). We saw that the parameter \( \gamma \) will have the same estimate regardless of the choice of \( \hat{\Pi} \) matrix used in the fitted model. Nevertheless, as long as there is a possibility of misclassification of the covariates (an error which is ignored in the following analyses), we cannot be certain that we are estimating \( \gamma^{*} \) even when \( \hat{\Pi} = \Pi \). If there is no possibility of covariate misclassification, the parameter \( \gamma \) in the above model can be replaced by \( \gamma^{*} \).

Using the theory developed in Chapter 3, we can conclude that \( \hat{\gamma} \) will not change for different values of \( \hat{\Pi} \). In addition, the predicted values and goodness-of-fit statistics will be invariant to changes in
This second approach to assessing the effect of misclassification will be followed in the remainder of this section. We will illustrate the use of models (1.16), (1.18), and (1.19) with logistic, Poisson, and regular unweighted least squares regression methods, respectively. Three data sets are presented, each of which is suspected to involve misclassification with respect to exposure. Each is analyzed using an appropriate regression procedure.

Five different choices for $\hat{\Pi}$ are considered, including $\hat{\Pi}=\hat{\Pi}^0$. The $\hat{\Pi}_i$ matrix corresponding to each of these choices is presented below.

$\hat{\Pi}_{1,1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

$\hat{\Pi}_{1,2} = \begin{bmatrix} 1 & .05 & .03 & .02 \\ 1 & .90 & .04 & .02 \\ 1 & .04 & .90 & .04 \\ 1 & .025 & .05 & .90 \end{bmatrix}$

$\hat{\Pi}_{1,3} = \begin{bmatrix} 1 & .098 & .001 & .001 \\ 1 & .90 & .04 & .03 \\ 1 & .04 & .90 & .03 \\ 1 & .04 & .05 & .90 \end{bmatrix}$

$\hat{\Pi}_{1,4} = \begin{bmatrix} 1 & .10 & .07 & .03 \\ 1 & .80 & .08 & .04 \\ 1 & .08 & .80 & .08 \\ 1 & .07 & .10 & .80 \end{bmatrix}$

$\hat{\Pi}_{1,5} = \begin{bmatrix} 1 & .01 & .001 & .001 \\ 1 & .80 & .10 & .08 \\ 1 & .08 & .80 & .08 \\ 1 & .07 & .08 & .80 \end{bmatrix}$

The first matrix above, $\hat{\Pi}_{1,1}$, is the matrix which assumes no misclassification (i.e., $\hat{\Pi}_{1,1}=\hat{\Pi}^0$). For the matrices $\hat{\Pi}_{1,2}$ and $\hat{\Pi}_{1,3}$, the
values of the \( \hat{\pi}_{ij} \)'s with \( i=j \) are each equal to .90. This implies that the probability of being correctly classified for both of these sets of misclassification probabilities is .90. The differences between the two matrices are in their off-diagonal elements. For the matrices \( \hat{\Pi}_{1,4} \) and \( \hat{\Pi}_{1,5} \), the values of the \( \hat{\pi}_{ij} \)'s with \( i=j \) are each equal to .80 for \( i,j=1,2,3 \). The value of the \( \pi_{0j} \) associated with \( \hat{\Pi}_{1,4} \) is also .80; the value of the \( \pi_{0j} \) associated with \( \hat{\Pi}_{1,5} \), however, is .988.

Finally, these choices of \( \hat{\Pi}_1 \) all assume a misclassification rate less than or equal to 20% (i.e., \( \hat{\pi}_{ij} \geq .80 \) for all \( i=j \)). From the discussion in Section 5.2, then, we can assume that models (1.16) and (1.18) will hold reasonably well. Therefore, we should not be concerned with the appropriateness of the two models in the following examples.

5.3.1 Logistic Regression Example

For this example, model (1.16) is assumed to hold. The classified data are given in Table 5.9. There are four exposure categories (i.e., \( k=3 \)) and one covariate which divides the data into two strata (i.e., \( S=2 \)). One covariate value is needed to represent the stratum effect (i.e., \( p=1 \)). Let us define \( v_s=(v_s) \) where \( v_1=0 \) and \( v_2=1 \). The \( \hat{\Pi} \) matrix for this example will be of the form \( \hat{\Pi} = \begin{bmatrix} \hat{\Pi}_1 \\ \hat{\Pi}_1 \end{bmatrix} \) since there are just two strata.
Table 5.9
Data for Example 5.3.1

<table>
<thead>
<tr>
<th></th>
<th>s=1</th>
<th>s=2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>E₀</td>
<td>2</td>
<td>E₀</td>
</tr>
<tr>
<td>E₁</td>
<td>5</td>
<td>E₁</td>
</tr>
<tr>
<td>E₂</td>
<td>15</td>
<td>E₂</td>
</tr>
<tr>
<td>E₃</td>
<td>4</td>
<td>E₃</td>
</tr>
</tbody>
</table>

The model being fit becomes

\[
\text{logit } \theta_{is} = \beta_\alpha + \sum_{j=1}^3 \pi_{ij} \beta_j + v_s \gamma_1.
\]

\( \beta_\alpha \) is the reference value for the group in the lowest exposure category and the first stratum. \( \beta_1, \beta_2, \) and \( \beta_3 \) are the effects for the second, third, and fourth exposure categories, respectively. \( \gamma_1 \) is the effect for the second stratum.

The five choices of \( \hat{\Pi}_1 \) presented earlier were used in five separate fitted models. The CATMAX procedure was used to produce both WLS and ML estimates of \( \beta \) and \( \gamma \). In addition, goodness of fit was measured for both types of estimation, and the predicted counts were calculated for ML estimation.

As was previously mentioned, \( \hat{\gamma}_{\text{wls}} \) and \( \hat{\gamma}_{\text{ml}} \), the WLS and ML estimates of \( \gamma_1 \), along with the WLS and ML goodness-of-fit (GOF) statistics, are invariant for all possibilities of \( \hat{\Pi}_1 \). These values are given below.

\[
\hat{\gamma}_{\text{wls}} = -0.0321  \quad Q_w = 0.4361, \quad p = 0.9327
\]

\[
\hat{\gamma}_{\text{ml}} = -0.0319  \quad Q_p = 0.4383, \quad p = 0.9322
\]

\[
Q_{\text{log}} = 0.4383, \quad p = 0.9322.
\]
where \( p \) represents the p-value associated with the value of the GOF statistic. The predicted counts are invariant as well.

The estimates of \( \beta \), however, are not invariant. Tables 5.10 and 5.11 contain these values for each of the \( \hat{\Pi}_i \) matrices for WLS and ML estimation procedures, respectively. Let us focus on the ML estimates. The values in the \( \hat{\Pi}_{1,1} \) column are the estimates of \( \beta \) obtained assuming the data are perfectly classified. The values contained in the remaining four columns are the estimates obtained assuming various degrees of misclassification. These can be compared to the left-hand column’s values to see how the estimates are affected when different sets of misclassification probabilities are assumed.

Table 5.10

<table>
<thead>
<tr>
<th>Structure of ( \hat{\Pi}_i ) matrix</th>
<th>( \hat{\Pi}_{1,1} )</th>
<th>( \hat{\Pi}_{1,2} )</th>
<th>( \hat{\Pi}_{1,3} )</th>
<th>( \hat{\Pi}_{1,4} )</th>
<th>( \hat{\Pi}_{1,5} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta}_1 )</td>
<td>.4860</td>
<td>.5563</td>
<td>.4749</td>
<td>.6413</td>
<td>.2539</td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>1.1810</td>
<td>1.3187</td>
<td>1.2830</td>
<td>1.4824</td>
<td>1.2540</td>
</tr>
<tr>
<td>( \hat{\beta}_3 )</td>
<td>1.7456</td>
<td>1.9695</td>
<td>1.9024</td>
<td>2.2343</td>
<td>2.0417</td>
</tr>
</tbody>
</table>

The estimate for \( \beta_3 \) must be smaller when \( \hat{\Pi}_1=\hat{\Pi}_{1,1} \) than when \( \hat{\Pi}_1 \) has any other value. We can see this by reviewing the theory in Section 4.5. There, we showed that when the model being fit is \( E(Y)=\hat{\Pi}\beta+\nu Y \), if \( \hat{\Pi}=\hat{\Pi}^0 \), \( \hat{\beta}_k^0 \) is biased towards zero. It can also be shown that \( \hat{\beta}_k^0 \) is less than \( \hat{\beta}_k^* \), where \( \hat{\beta}_k^* \) is the estimate obtained when the correct \( \Pi \) matrix is used in the fitted model. In fact, \( \hat{\beta}_k^0 \)
is less than \( \hat{\beta}_k \), where \( \hat{\beta}_k \) is the estimate obtained when any \( \hat{\Pi} \) matrix is used in the fitted model. Intuitively, this is reasonable considering that any \( \hat{\Pi} \) could be the true \( \Pi \) matrix.

Therefore, when the model being fit is \( E(Y) = \Pi \beta + \nu_Y \), as it is in this section, the estimate \( \hat{\beta}_k \) obtained when \( \hat{\Pi} = \hat{\Pi}^0 \) will be less than the estimate obtained when any other structure of \( \hat{\Pi} \) is assumed. Also, recall that nothing could be said of the direction of the bias in estimating \( \beta_1 \) and \( \beta_2 \). It follows, then, that the estimates of these parameters obtained when \( \hat{\Pi} = \hat{\Pi}^0 \) will not necessarily be smaller than those obtained when \( \hat{\Pi} \) takes on another structure.

**Table 5.11**

ML Estimates \( \hat{\beta}_j, j=1,2,3 \), (and Estimated Standard Errors), for Example 5.3.1

<table>
<thead>
<tr>
<th>Structure of ( \hat{\Pi}_1 ) matrix</th>
<th>( \hat{\Pi}_{1,1} )</th>
<th>( \hat{\Pi}_{1,2} )</th>
<th>( \hat{\Pi}_{1,3} )</th>
<th>( \hat{\Pi}_{1,4} )</th>
<th>( \hat{\Pi}_{1,5} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>.4929</td>
<td>.5643</td>
<td>.4832</td>
<td>.6510</td>
<td>.2622</td>
</tr>
<tr>
<td></td>
<td>(.6078)</td>
<td>(.7092)</td>
<td>(.7149)</td>
<td>(.8517)</td>
<td>(.6736)</td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>1.1870</td>
<td>1.3257</td>
<td>1.2903</td>
<td>1.4911</td>
<td>1.2613</td>
</tr>
<tr>
<td></td>
<td>(.5674)</td>
<td>(.6462)</td>
<td>(.6496)</td>
<td>(.7543)</td>
<td>(.6241)</td>
</tr>
<tr>
<td>( \hat{\beta}_3 )</td>
<td>1.7427</td>
<td>1.9662</td>
<td>1.8993</td>
<td>2.2304</td>
<td>2.0367</td>
</tr>
<tr>
<td></td>
<td>(.6520)</td>
<td>(.7441)</td>
<td>(.7303)</td>
<td>(.8487)</td>
<td>(.7496)</td>
</tr>
</tbody>
</table>

The \( \hat{\text{OR}}_j \)'s corresponding to the ML estimates are given in Table 5.12, along with large-sample 95% confidence intervals. These intervals were calculated using the formula

\[
\{ \exp [ \hat{\beta}_j \pm 1.96 \text{ (estimated standard error of } \hat{\beta}_j)] \}. \]

The values of
Table 5.11

Estimates $\hat{OR}_j, j=1,2,3$, (and Corresponding 95% Confidence Intervals) for Example 5.3.1

<table>
<thead>
<tr>
<th>Structure of $\hat{\Pi}_i$ matrix</th>
<th>$\hat{\Pi}_{1,1}$</th>
<th>$\hat{\Pi}_{1,2}$</th>
<th>$\hat{\Pi}_{1,3}$</th>
<th>$\hat{\Pi}_{1,4}$</th>
<th>$\hat{\Pi}_{1,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{OR}_1$</td>
<td>1.64</td>
<td>1.76</td>
<td>1.62</td>
<td>1.92</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>(.50, 5.39)</td>
<td>(.44, 7.06)</td>
<td>(.40, 10.18)</td>
<td>(.36, 10.18)</td>
<td>(.35, 4.87)</td>
</tr>
<tr>
<td>$\hat{OR}_2$</td>
<td>3.28</td>
<td>3.77</td>
<td>3.63</td>
<td>4.44</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>(1.08, 9.97)</td>
<td>(1.06, 13.36)</td>
<td>(1.02, 12.98)</td>
<td>(1.01, 19.48)</td>
<td>(1.04, 12.00)</td>
</tr>
<tr>
<td>$\hat{OR}_3$</td>
<td>5.71</td>
<td>7.14</td>
<td>6.68</td>
<td>9.30</td>
<td>7.67</td>
</tr>
<tr>
<td></td>
<td>(1.60, 20.50)</td>
<td>(1.66, 30.71)</td>
<td>(1.60, 27.96)</td>
<td>(1.76, 49.10)</td>
<td>(1.76, 33.31)</td>
</tr>
</tbody>
</table>
\( \hat{OR}_1 \) do not vary a lot over the five columns. They range from 1.30 to 1.92. Likewise, the values of \( \hat{OR}_2 \) range from 3.28 to 4.44, not a large variation. The values of \( \hat{OR}_3 \) range from 5.71 to 9.30. One conclusion which can be drawn is that \( \hat{\beta}_3^* > 5.71 \), since \( \hat{\beta}_k^0 \) is smaller than the corresponding estimate for any other choice of \( \hat{\Pi}_1 \).

For each \( \hat{OR}_j \), the corresponding confidence intervals have lower bounds which vary little between values of \( \hat{\Pi}_1 \). The upper bounds, however, vary quite a lot. Also, notice that the confidence intervals for each \( \hat{OR}_j \) either all include the null value of 1, or all do not include the null value of 1.

5.3.2 Poisson Regression Example

In this example, we assume model (1.18) holds. The classified data are given in Table 5.13. These are hypothetical data which could reflect the incidence of disease (D) arising out of a total amount of population-time at risk (L). Again, there are four exposure categories. In addition, there are six strata defined by six levels of one covariate. The matrices \( \hat{\Pi} \) and \( V_i \) have the following forms:

\[
\hat{\Pi} = \begin{bmatrix}
\hat{\Pi}_1 \\
\hat{\Pi}_2 \\
\hat{\Pi}_3 \\
\hat{\Pi}_4 \\
\hat{\Pi}_5 \\
\hat{\Pi}_6
\end{bmatrix}
\quad \text{and} \quad
V_i = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

where \( \hat{\Pi}_i \) can be one of the choices given earlier.

The model to be fit becomes
\[
\ln \lambda_{is} = \beta_\alpha + \sum_{j=1}^{3} \pi_{ij} \beta_j + v_s' \gamma
\]

where \( \beta_\alpha \) is the reference value for the group in the lowest exposure category and the first stratum. \( \beta_1, \beta_2, \) and \( \beta_3 \) are the effects for the second, third, and fourth exposure categories, respectively. \( \gamma_1, \gamma_2, \ldots, \gamma_5 \) are the effects for stratum 2, 3, \ldots, 6, respectively, so that \( \gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_5) \).

**Table 5.13**

Data for Example 5.3.2

<table>
<thead>
<tr>
<th></th>
<th>s=1</th>
<th>s=2</th>
<th>s=3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>L</td>
<td>D</td>
</tr>
<tr>
<td>E_0</td>
<td>186</td>
<td>22013</td>
<td>1648</td>
</tr>
<tr>
<td>E_1</td>
<td>23</td>
<td>1974</td>
<td>530</td>
</tr>
<tr>
<td>E_2</td>
<td>2</td>
<td>133</td>
<td>128</td>
</tr>
<tr>
<td>E_3</td>
<td>1249</td>
<td>97469</td>
<td>3692</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>s=4</th>
<th>s=5</th>
<th>s=6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>L</td>
<td>D</td>
</tr>
<tr>
<td>E_0</td>
<td>764</td>
<td>67290</td>
<td>356</td>
</tr>
<tr>
<td>E_1</td>
<td>769</td>
<td>60348</td>
<td>453</td>
</tr>
<tr>
<td>E_2</td>
<td>475</td>
<td>32437</td>
<td>340</td>
</tr>
<tr>
<td>E_3</td>
<td>662</td>
<td>35724</td>
<td>308</td>
</tr>
</tbody>
</table>

The Poisson option of the CATMAX procedure was used to obtain WLS and ML estimates. Again, the values of \( \hat{\gamma}_{wls} \) and \( \hat{\gamma}_{ml} \) are common to all \( \hat{\Pi} \) matrices used in the fitted model. These values are given below along with the values of the common goodness-of-fit statistics for WLS and ML estimation.
\[ \hat{\gamma}_{\text{wls}} = \begin{bmatrix} .0319 \\ .1211 \\ .3337 \\ .7276 \\ 1.0896 \end{bmatrix} \quad Q_w = 15.1497, \ p = .4414 \]

\[ \hat{\gamma}_{\text{ml}} = \begin{bmatrix} .0323 \\ .1214 \\ .3339 \\ .7270 \\ 1.0857 \end{bmatrix} \quad Q_p = 15.1726, \ p = .4391 \quad Q_{\log} = 15.1913, \ p = .4377 \]

In addition, the standard errors of the \( \hat{\gamma} \)'s are also invariant.

As in the previous example, the \( \hat{\beta} \)'s for WLS and ML methods were calculated assuming various structures for the \( \hat{\Pi}_i \) matrix. Table 5.14 contains the resulting ML estimates and their standard errors. Table 5.15 contains the corresponding estimates of the IDR\(_j\)'s and 95% confidence intervals.
Table 5.14
ML Estimates $\hat{\beta}_j$, $j=1,2,3$ (and Estimated Standard Errors) for Example 5.3.2

<table>
<thead>
<tr>
<th>Structure of $\hat{\Pi}_1$ matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Pi}_{1,1}$</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td>(.0241)</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
</tr>
<tr>
<td>(.0308)</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
</tr>
<tr>
<td>(.0189)</td>
</tr>
</tbody>
</table>

Notice that although the same $\hat{\Pi}_1$'s matrices are used here as in the previous example, the ranges in this example are smaller. In addition, the upper limits of the confidence intervals for the IDR. j's are much more stable than those in Example 5.3.1. This indicates that the analyses of certain data sets are more sensitive to changes in misclassification probabilities than are others.

A series of tests were performed to relate the theory developed in Section 4.7 to this example. The null hypothesis $H_0: \beta_j = 0$ was tested for $j=1,2,...,k$. Table 5.9 contains the $Q_{wc}$ values associated with each test under each assumed $\hat{\Pi}_1$ matrix. According to the theory presented in Section 4.7, these values should vary, which, in fact, they do. In addition, when the hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ is tested, the resulting $Q_{wc}$ statistic has the value $Q_{wc} = 578.756$ for each $\hat{\Pi}_1$, as the theory dictates should happen.
### Table 5.15

**ML Estimates IDR_{j, j=1,2,3}, (and Corresponding 95% Confidence Intervals) for Example 5.3.2**

<table>
<thead>
<tr>
<th>Structure of $\hat{\Pi}_1$ matrix</th>
<th>$\hat{\Pi}_{1,1}$</th>
<th>$\hat{\Pi}_{1,2}$</th>
<th>$\hat{\Pi}_{1,3}$</th>
<th>$\hat{\Pi}_{1,4}$</th>
<th>$\hat{\Pi}_{1,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDR_{1}</td>
<td>1.12</td>
<td>1.14</td>
<td>1.12</td>
<td>1.16</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(1.07, 1.17)</td>
<td>(1.08, 1.20)</td>
<td>(1.06, 1.19)</td>
<td>(1.09, 1.24)</td>
<td>(1.01, 1.13)</td>
</tr>
<tr>
<td>IDR_{2}</td>
<td>1.23</td>
<td>1.26</td>
<td>1.25</td>
<td>1.29</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>(1.16, 1.31)</td>
<td>(1.17, 1.35)</td>
<td>(1.17, 1.34)</td>
<td>(1.18, 1.40)</td>
<td>(1.14, 1.32)</td>
</tr>
<tr>
<td>IDR_{3}</td>
<td>1.55</td>
<td>1.65</td>
<td>1.63</td>
<td>1.77</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>(1.50, 1.61)</td>
<td>(1.58, 1.72)</td>
<td>(1.56, 1.70)</td>
<td>(1.68, 1.85)</td>
<td>(1.62, 1.77)</td>
</tr>
</tbody>
</table>
Table 5.16
$Q_{wc}'s$ for Testing $H_0: \beta_j=0, j=1,2,3$

<table>
<thead>
<tr>
<th>$\beta_j$</th>
<th>$\hat{\Pi}_{1,1}$</th>
<th>$\hat{\Pi}_{1,2}$</th>
<th>$\hat{\Pi}_{1,3}$</th>
<th>$\hat{\Pi}_{1,4}$</th>
<th>$\hat{\Pi}_{1,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>21.93</td>
<td>21.25</td>
<td>14.90</td>
<td>19.26</td>
<td>4.98</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>46.01</td>
<td>42.36</td>
<td>41.79</td>
<td>35.71</td>
<td>28.83</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>542.72</td>
<td>533.94</td>
<td>519.28</td>
<td>529.44</td>
<td>553.26</td>
</tr>
</tbody>
</table>

Although the values of $Q_{wc}$ may seem to vary greatly, the conclusions reached using the different $\hat{\Pi}_1$ matrices are the same. All five sets of results indicate that $H_0: \beta_1=0$ should be rejected, along with $H_0: \beta_2=0$ and $H_0: \beta_3=0$. This is in accordance with the conclusions reached after examining the confidence intervals in Table 5.15. None of the fifteen intervals contains the null value of 1. In particular, it would be comforting to know that the analysis results are fairly robust to different specifications for the $\hat{\Pi}_1$ matrix.

5.3.3 Least Squares Regression Example

In this example, we assume the following model holds:

$E(Y_{is}) = \sum_{j=0}^{k} \pi_{ij} E(Y_{js})$ where $Y$ is a continuous variable, in this case, blood pressure. The classified data are given in Table 5.17. There are four exposure categories (i.e., $k=3$). The covariates involved are continuous age and weight variables. Therefore, $p=2$. Each $s$ subscript refers to a specific subject within a classified exposure category. There are three subjects classified as unexposed, two
subjects classified into each of the first and second exposed categories, and one subject classified into the third (or highest) exposed category. Since the numbers of subjects classified into each exposure category are not all equal, the discussion in Section 4.9 pertains to this situation.

The model to be fit is $E(Y_{is}) = \beta_\alpha + \sum_{j=1}^{3} \pi_{ij} \beta_j + v_s \gamma_i$. $\beta_\alpha$ is the intercept term. $\beta_1$, $\beta_2$ and $\beta_3$ are the effects for the first, second and third exposed categories, respectively. $v_s'=(v_{s1}, v_{s2})$ where $v_{s1}$ is the value of the age variable and $v_{s2}$ is the value of the weight variable. $\gamma_1$ is the age effect and $\gamma_2$ is the weight effect.

### Table 5.17
Data for Example 5.3.3

<table>
<thead>
<tr>
<th></th>
<th>blood pressure</th>
<th>age</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_0$</td>
<td>s=1</td>
<td>110</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>s=2</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>s=3</td>
<td>98</td>
<td>27</td>
</tr>
<tr>
<td>$E_1$</td>
<td>s=1</td>
<td>110</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>s=2</td>
<td>114</td>
<td>31</td>
</tr>
<tr>
<td>$E_2$</td>
<td>s=1</td>
<td>112</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>s=2</td>
<td>118</td>
<td>28</td>
</tr>
<tr>
<td>$E_3$</td>
<td>s=1</td>
<td>113</td>
<td>36</td>
</tr>
</tbody>
</table>

When the $Y_{is}$'s are given in the order $Y=(Y_{01}, Y_{02}, Y_{03}, Y_{11}, Y_{12}, Y_{21}, Y_{22}, Y_{31})'$, the $\Pi$ matrix will have the form.
\[
\hat{\Pi} = \begin{bmatrix}
1 & \hat{\pi}_{01} & \hat{\pi}_{02} & \hat{\pi}_{03} \\
1 & \hat{\pi}_{11} & \hat{\pi}_{12} & \hat{\pi}_{13} \\
1 & \hat{\pi}_{11} & \hat{\pi}_{12} & \hat{\pi}_{13} \\
1 & \hat{\pi}_{11} & \hat{\pi}_{12} & \hat{\pi}_{13} \\
1 & \hat{\pi}_{11} & \hat{\pi}_{12} & \hat{\pi}_{13} \\
1 & \hat{\pi}_{21} & \hat{\pi}_{22} & \hat{\pi}_{23} \\
1 & \hat{\pi}_{21} & \hat{\pi}_{22} & \hat{\pi}_{23} \\
1 & \hat{\pi}_{31} & \hat{\pi}_{32} & \hat{\pi}_{33}
\end{bmatrix}.
\]

For this example, the same \( \hat{\Pi}_i \)'s used earlier are considered. The estimates \( \hat{\beta}_j \), \( j=1,2,3 \), and their estimated standard errors are given in Table 5.18. As expected the values of \( \hat{\gamma}_1 \) and \( \hat{\gamma}_2 \) are the same for all five choices of \( \hat{\Pi}_i \). In addition, the estimated standard errors of the \( \hat{\gamma}'s \), \( R^2 \), and the F statistic are invariant to choices of \( \hat{\Pi}_i \). This is consistent with the theory for LS estimation, presented in Section 4.9, which deals with situations in which there are unequal numbers of subjects in the classified exposure categories.

**Table 5.18**

| LS Estimates \( \hat{\beta}_j \), \( j=1,2,3 \) (and Estimated Standard Errors) for Example 5.3.3 |
|---|---|---|---|---|---|
| Structure of \( \hat{\Pi}_i \) matrix | \( \hat{\Pi}_{1,1} \) | \( \hat{\Pi}_{1,2} \) | \( \hat{\Pi}_{1,3} \) | \( \hat{\Pi}_{1,4} \) | \( \hat{\Pi}_{1,5} \) |
| \( \hat{\beta}_1 \) | 29.355 | 33.992 | 32.055 | 40.117 | 25.144 |
| (10.100) | (11.706) | (11.381) | (13.956) | (9.732) |
| \( \hat{\beta}_2 \) | 41.149 | 46.172 | 45.769 | 52.331 | 42.711 |
| (14.123) | (16.037) | (15.954) | (18.620) | (15.610) |
| \( \hat{\beta}_3 \) | 58.110 | 65.950 | 64.212 | 74.991 | 66.617 |
The estimated values of $\hat{\gamma}$ (and their standard errors), of $R^2$, and of the F statistic are given below. The F statistic used in this analysis tests that all the coefficients except the intercept term, $\beta_1$, are zero.

$\hat{\gamma}_1 = -1.410$ \hspace{1cm} $R^2 = .9626$ 
$(.642)$

$\hat{\gamma}_2 = .958$ \hspace{1cm} $F = 10.301, \ p = .0908$ 
$(.546)$

By comparing the estimates in Tables 5.11, 5.14, and 5.18, we can see that the variation in absolute terms among values in the last table is much greater. In relative terms, however, the degrees of variation are more or less the same among the three tables. In fact, relative to the estimates in the $\hat{\Pi}_{1,1}$ column, the range for $\hat{\beta}_1$ is less in Table 5.18 than in Tables 5.11 and 5.14.

Certain patterns among the estimates are similar across the three tables. This indicates that estimators from the three types of regression procedures are affected by varying degrees of misclassification in similar ways. For instance, compare the first rows of Tables 5.11, 5.14, and 5.18. The lowest value in each of these rows corresponds to the $\hat{\Pi}_{1,5}$ matrix; the highest value in each corresponds to the $\hat{\Pi}_{1,4}$ matrix. In fact, the highest value in each row for all three tables corresponds to the $\hat{\Pi}_{1,4}$ matrix. The second highest value in the second row of all three tables corresponds to the $\hat{\Pi}_{1,2}$ matrix, and in the third row to the $\hat{\Pi}_{1,5}$ matrix. From this we can see that the type of regression procedure used does not affect
certain relationships among the estimates resulting from the use of various $\hat{\Pi}$ matrices.

5.4.3 Real Life Example

As a final example, we will investigate the effect of misclassification of exposure on a data set given by Frome (1983). The classified data appear in Table 5.19. These data are a subset of the original data, including only the smokers.

The exposure variable is "cigarettes per day". This variable is considered to involve potential misclassification error. There are six levels of the variable so that $k=5$. The covariate is "years of smoking", which will not be considered to involve misclassification. There are nine levels of this variable, each of which represents a stratum, so that $S=9$. Therefore, there are a total of 54 $(i,s)$ cells. The values in the table represent the number of man-years at risk and the observed number of lung cancer deaths (in parentheses) for each $(i,s)$ cell.

The variable $d_j$, $j=0,1,2,3,4,5$, represents the dosage level variable for the $j$-th true exposure category. The variable $t_s$, $s=1,2,...,9$, is equal to the midpoint years of smoking in stratum $s$ divided by 42.5. Functions of these variables, namely $\ln d_j$ and $\ln t_s$, are used later as sets of score values for these two variables.

Two different analyses are performed on the data. First, "cigarettes per day" is treated as a nominal variable. The model used for this analysis includes five separate exposure effects, one for each exposure level above the lowest. The second analysis treats
Table 5.19
Man-Years at Risk (and Observed Number of Lung Cancer Deaths)

<table>
<thead>
<tr>
<th>Years of Smoking</th>
<th>Cigarettes per day</th>
<th>(E_0)</th>
<th>(E_1)</th>
<th>(E_2)</th>
<th>(E_3)</th>
<th>(E_4)</th>
<th>(E_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_j</td>
<td></td>
<td>5.2</td>
<td>11.2</td>
<td>15.9</td>
<td>20.4</td>
<td>27.4</td>
<td>40.8</td>
</tr>
<tr>
<td>15-19</td>
<td></td>
<td>3,121</td>
<td>3,544</td>
<td>4,317</td>
<td>5,683</td>
<td>3,042</td>
<td>670</td>
</tr>
<tr>
<td>20-24</td>
<td></td>
<td>2,937</td>
<td>3,286</td>
<td>4,214</td>
<td>6,385</td>
<td>4,050</td>
<td>1,166</td>
</tr>
<tr>
<td>25-29</td>
<td></td>
<td>2,288</td>
<td>2,546</td>
<td>3,185</td>
<td>5,483</td>
<td>4,290</td>
<td>1,482</td>
</tr>
<tr>
<td>30-34</td>
<td></td>
<td>2,015</td>
<td>2,219</td>
<td>2,560</td>
<td>4,687</td>
<td>4,268</td>
<td>1,580</td>
</tr>
<tr>
<td>35-39</td>
<td></td>
<td>1,648</td>
<td>1,826</td>
<td>1,893</td>
<td>3,646</td>
<td>3,529</td>
<td>1,336</td>
</tr>
<tr>
<td>40-44</td>
<td></td>
<td>1,310</td>
<td>1,386</td>
<td>1,334</td>
<td>2,411</td>
<td>2,424</td>
<td>924</td>
</tr>
<tr>
<td>45-49</td>
<td></td>
<td>927</td>
<td>988</td>
<td>849</td>
<td>1,567</td>
<td>1,409</td>
<td>556</td>
</tr>
<tr>
<td>50-54</td>
<td></td>
<td>710</td>
<td>684</td>
<td>470</td>
<td>857</td>
<td>663</td>
<td>255</td>
</tr>
<tr>
<td>55-59</td>
<td></td>
<td>606</td>
<td>449</td>
<td>280</td>
<td>416</td>
<td>284</td>
<td>104</td>
</tr>
</tbody>
</table>
exposure as an ordinal variable. Here, a dose-response model is fit to the data. In both analyses, the covariate is represented by $\ln t_s$ and is treated as a continuous variable. In the second analysis, "cigarettes per day" is represented by $\ln d_j$, a continuous variable.

Under each analysis, an appropriate misclassification model is used to analyze the data assuming various degrees of misclassification of exposure. Model (1.16) is used for the first analysis; model (2.10) is used for the second. Five different sets of misclassification probabilities are assumed. The same sets are used for both analyses. The values of the $\hat{\pi}_{i,j}$'s in the five sets are presented below.

The first set corresponds to no misclassification. The third and fourth sets could reflect situations in which it is more likely for subjects to be misclassified into lower exposure categories than into higher exposure categories. The pattern of $\hat{\pi}_{i,j}$'s in Set #2 is such that, for each classified exposure category, the probability of truly belonging in an adjacent category is the same (.04). Further, the probability of truly belonging in a category which is two categories away from the classified category is the same for all classified categories (.01). The $\hat{\pi}_{i,j}$'s in Set #5 reflect a situation in which those subjects classified into the lowest exposure category are least likely to be misclassified.
<table>
<thead>
<tr>
<th>i=0</th>
<th>j=0</th>
<th>j=1</th>
<th>j=2</th>
<th>j=3</th>
<th>j=4</th>
<th>j=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i=0</th>
<th>j=0</th>
<th>j=1</th>
<th>j=2</th>
<th>j=3</th>
<th>j=4</th>
<th>j=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.95</td>
<td>.04</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.04</td>
<td>.91</td>
<td>.04</td>
<td>.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td>.04</td>
<td>.90</td>
<td>.04</td>
<td>.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i=3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.01</td>
<td>.04</td>
<td>.90</td>
<td>.04</td>
<td>.01</td>
<td>0</td>
</tr>
<tr>
<td>i=4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.04</td>
<td>.91</td>
<td>.04</td>
<td>0</td>
</tr>
<tr>
<td>i=5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>.04</td>
<td>.95</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i=0</th>
<th>j=0</th>
<th>j=1</th>
<th>j=2</th>
<th>j=3</th>
<th>j=4</th>
<th>j=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.85</td>
<td>.10</td>
<td>.03</td>
<td>.015</td>
<td>.003</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>i=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td>.87</td>
<td>.07</td>
<td>.03</td>
<td>.014</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>i=2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.002</td>
<td>.008</td>
<td>.88</td>
<td>.07</td>
<td>.03</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>i=3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.001</td>
<td>.001</td>
<td>.008</td>
<td>.89</td>
<td>.07</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>i=4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0003</td>
<td>.0007</td>
<td>.001</td>
<td>.008</td>
<td>.90</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>i=5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0003</td>
<td>.0007</td>
<td>.001</td>
<td>.003</td>
<td>.008</td>
<td>.987</td>
<td></td>
</tr>
</tbody>
</table>

163
5.4.1 Analysis Assuming Nominal Exposure Categories

The data in Table 5.19 are analyzed here assuming nominal exposure categories and ordinal covariates levels. The variable $\ln t_s$, $s=1,2,...,9$, is used to represent the levels of the covariate "years of smoking". Under assumed degrees of misclassification, model (1.16) becomes

$$\logit \theta_{is}(v_s) = \beta_0 + \sum_{j=1}^{5} \pi_{ij} \beta_j + v_s \gamma$$

where $\beta_0$ is the effect for the $j=0$ exposure category, $\beta_j$ is the effect for the $j$-th exposure category, $j=1,2,3,4,5$; $v_s = \ln t_s$, and $\gamma$ is the
linear effect of the covariate.

The $\hat{H}$ matrix is a series of nine vertically concatenated $\hat{H}_i$ matrices, where $\hat{H}_i$ is an estimate of $H_i$, defined in Section 1.10. Five different structures of $\hat{H}_i$ are assumed, each corresponding to one of the five sets of $\hat{\pi}_{ij}$'s given earlier. The resulting ML estimates of the $\beta_j$'s, $j=1,2,3,4,5$, and their estimated standard errors are given in Table 5.20. The corresponding estimates of the IDR $j$'s and their 95% large-sample confidence intervals are given in Table 5.21. As expected, $\hat{\gamma}$ and its estimated standard error are invariant to choices of $\hat{H}$. These values, along with the GOF statistics and corresponding p-values, are given below:

\[
\hat{\gamma} = 4.5023 \quad \text{Estimated standard error of } \hat{\gamma} = .3356
\]

\[
Q_p = 39.7807 \quad p=.7633
\]

\[
Q_{log} = 47.1459 \quad p=.4666
\]

Inspection of Table 5.20 shows that the estimates $\hat{\beta}_j$, $j=1,2,3,4,5$, corresponding to Sets #1 through #4, and their estimated standard errors, increase monotonically. Not all of the estimates $\hat{\beta}_j$, $j=1,2,3,4,5$, corresponding to Set #5, however, are larger than those corresponding to Set #4. The estimate $\hat{\beta}_1$ for Set #5 falls between the estimates of $\beta_1$ for Sets #2 and #3; $\hat{\beta}_2$ for Set #5 falls between the $\hat{\beta}_2$'s for Sets #3 and #4. Further, none of the estimated standard errors for Set #5 is larger than for Set #4.

These relationships illustrate the point that the relationships among the $\hat{\beta}$'s corresponding to different sets of misclassification probabilities are
Table 5.20
ML Estimates $\hat{\beta}_j$ (and Estimated Standard Errors), $j=1,2,3,4,5$, for Example 5.4.1

<table>
<thead>
<tr>
<th>Assumed Set of $\hat{\pi}_{ij}$'s</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8780</td>
<td>0.9502</td>
<td>1.0005</td>
<td>1.1128</td>
<td>0.9535</td>
</tr>
<tr>
<td></td>
<td>(0.4880)</td>
<td>(0.5444)</td>
<td>(0.5960)</td>
<td>(0.6533)</td>
<td>(0.5927)</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>1.0924</td>
<td>1.1268</td>
<td>1.1861</td>
<td>1.2497</td>
</tr>
<tr>
<td></td>
<td>(0.4838)</td>
<td>(0.5235)</td>
<td>(0.5703)</td>
<td>(0.6052)</td>
<td>(0.5818)</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_3$</td>
<td>1.6841</td>
<td>1.7511</td>
<td>1.8292</td>
<td>1.9233</td>
</tr>
<tr>
<td></td>
<td>(0.4338)</td>
<td>(0.4604)</td>
<td>(0.5110)</td>
<td>(0.5442)</td>
<td>(0.5265)</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_4$</td>
<td>1.9072</td>
<td>1.9518</td>
<td>2.0363</td>
<td>2.1167</td>
</tr>
<tr>
<td></td>
<td>(0.4321)</td>
<td>(0.4581)</td>
<td>(0.5078)</td>
<td>(0.5398)</td>
<td>(0.5254)</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_5$</td>
<td>2.3976</td>
<td>2.4750</td>
<td>2.5818</td>
<td>2.6573</td>
</tr>
<tr>
<td></td>
<td>(0.4456)</td>
<td>(0.4702)</td>
<td>(0.5156)</td>
<td>(0.5476)</td>
<td>(0.5433)</td>
</tr>
</tbody>
</table>

...influenced by more than the probabilities of correct classification. As we saw in Section 4.9, the relationships between estimates of $\beta$ corresponding to different sets of $\hat{\pi}_{ij}$'s do not depend on the relationships between the values of the $\hat{\pi}_{ij}$'s themselves, but on the relationships between the differences ($\hat{\pi}_{ij} - \hat{\pi}_{0j}$), $i,j=1,2,...,5$. In other words, we ask not whether a certain misclassification probability is larger in Set #2 than in Set #3; rather, we ask whether that misclassification probability minus the corresponding misclassification probability among the lowest classified exposure category is larger in Set #2 than in Set #3. The relationship of $\hat{\beta}$ for Set #5 and the $\hat{\beta}$'s for Sets #1 through #4 shows us that not all the estimates from one set will
Table 5.21
Estimates $\hat{IDR}_j$ (and Corresponding 95% Confidence Intervals)
$j=1,2,3,4,5$ for Example 5.4.1

<table>
<thead>
<tr>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{IDR}_1$</td>
<td>2.4 (.9, 6.3)</td>
<td>2.6 (.9, 7.5)</td>
<td>2.7 (.9, 8.8)</td>
<td>3.0 (.9, 11.0)</td>
</tr>
<tr>
<td>$\hat{IDR}_2$</td>
<td>3.0 (1.2, 7.7)</td>
<td>3.1 (1.1, 8.6)</td>
<td>3.3 (1.1, 10.0)</td>
<td>3.5 (1.1, 11.4)</td>
</tr>
<tr>
<td>$\hat{IDR}_3$</td>
<td>5.4 (2.3, 12.6)</td>
<td>5.8 (2.3, 14.2)</td>
<td>6.2 (2.3, 17.0)</td>
<td>6.8 (2.4, 19.9)</td>
</tr>
<tr>
<td>$\hat{IDR}_4$</td>
<td>6.7 (2.9, 15.7)</td>
<td>7.0 (2.9, 17.3)</td>
<td>7.7 (2.8, 20.7)</td>
<td>8.3 (2.9, 23.9)</td>
</tr>
<tr>
<td>$\hat{IDR}_5$</td>
<td>11.0 (4.6, 26.3)</td>
<td>11.9 (4.7, 29.9)</td>
<td>13.2 (4.8, 36.3)</td>
<td>14.3 (4.9, 41.7)</td>
</tr>
</tbody>
</table>
not necessarily be larger or smaller than the corresponding estimates from another set.

From Table 5.21, we can see the monotonic increase of the $\hat{\beta}_j$'s for Sets #1 through #4 reflected in the monotonic increase of the IDR$_j$'s for these sets. Notice, also, that the values of these estimates do not vary greatly among the first four sets of misclassification probabilities. The estimates resulting from use of the fifth set, however, are quite large compared to the values in the column for Set #1 for $j=3, 4, 5$

The upper bounds of the confidence intervals vary greatly, especially as $j$ increases. This is not unexpected since exponentiation is used in calculating the confidence intervals. The lower bounds are relatively stable, especially for Sets #1 through #4. The intervals across the sets for each $j$, however, are consistent with respect to the inclusion of the null value of 1. They either all contain 1 or none contains 1.

In addition, the hypothesis $H_0$: $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5$ was tested for each assumed set of $\hat{\pi}_{ij}$'s. As was mentioned in Section 4.8, the value of $Q_{wc}$ used to test this hypothesis varies with the choice of $\hat{\Pi}_i$. The values for the $Q_{wc}$'s and their corresponding p-values are given in Table 5.22.
Table 5.22

Values of $Q_{wc}$ (and P-Values) Based on Testing $H_0: \beta_1=\beta_2=\beta_3=\beta_4=\beta_5$

for Example 5.4.1

<table>
<thead>
<tr>
<th>Assumed Set of $\hat{\pi}_{ij}$s</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_{wc}$</td>
<td>31.91</td>
<td>29.68</td>
<td>31.32</td>
<td>30.82</td>
<td>32.96</td>
</tr>
<tr>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td></td>
</tr>
</tbody>
</table>

Although these values do vary, the degree of variation is small. Also, they all lead to the same conclusion: reject $H_0$.

Finally, let us draw some conclusions based on the results from using the five sets of $\hat{\pi}_{ij}$'s. Most importantly, we can see that, on the whole, the results reached when misclassification is ignored still hold when misclassification is not ignored. Many of the conclusions reached when using Set #1 are also reached when using Sets #2 through #5. The values of $Q_p$ and of $Q_{log}$ are invariant, indicating that the fit of the model is good for all of the $\hat{\pi}_{ij}$ sets. Also, the values of $Q_{wc}$ are nearly invariant, so that the conclusion of inequality among the true exposure effects is supported for all five sets.

In order to determine final estimates for the $\beta_j$'s and their estimated standard errors, we must review the five sets of misclassification probabilities. In general, when one is dealing with a situation in which there is potential misclassification, an important beginning step is to examine the exposure variable itself for clues.
concerning patterns and degrees of misclassification. For this example, let us make the following assumption. It is reasonable to suspect that subjects are likely to report that they smoke less than they really do and unlikely to report that they smoke more than they really do. In other words, those classified into low exposure categories are often likely to belong truly in higher categories. The converse, however, is not true; a subject who admits to smoking a large number of cigarettes probably does. As a result, the patterns in Sets #3 and #4 would seem to best reflect the true misclassification patterns.

Therefore, we suspect that the estimates resulting from the use of Sets #3 and #4 may be closest to the true estimates. Since the final estimate of $\beta$ will be a biased estimate of $\beta^*$, there is no need to worry over about obtaining a unique value for $\hat{\beta}$. Estimates from either Set #3 or #4 could be used. A final analysis would include these estimates, as well as the estimates obtained when misclassification is ignored.

5.4.2 Analysis Assuming Ordinal Exposure Categories

The data from Table 5.19 will now be analyzed assuming ordinal levels for the exposure variable as well as for the covariate. Again, the variable ln $t_s$, $s=1,2,...,9$, is used to represent the levels of the covariate, "years of smoking". The variable ln $d_j$ is used to represent the levels of the exposure variable, "cigarettes per day".

Model (2.10) can be applied in this situation. The score values,
c_j, j=0,1,...,k, (defined in Section 2.7) are equal to the values ln d_j, j=0,1,...,k, in this example. Under assumed sets of misclassification probabilities, the model becomes

$$\text{logit} \theta_{is} (v_s) = \beta_{\alpha} + \tau \sum_{j=0}^{5} \hat{\pi}_{ij} (\ln d_j) + v_s \gamma,$$

where $\beta_{\alpha}$ is the intercept term and $\tau$ is the linear effect for number of cigarettes per day; $v_s = \ln t_s$, and $\gamma$ is the linear effect for the years of smoking.

The $\hat{\Pi}$ matrix is a series of nine vertically concatenated $\hat{\Pi}_i$ matrices where

$$\hat{\Pi}_i = \begin{bmatrix}
1 & \sum_{j=0}^{5} \hat{\pi}_{0j} (\ln d_j) \\
1 & \sum_{j=0}^{5} \hat{\pi}_{1j} (\ln d_j) \\
1 & \sum_{j=0}^{5} \hat{\pi}_{2j} (\ln d_j) \\
1 & \sum_{j=0}^{5} \hat{\pi}_{3j} (\ln d_j) \\
1 & \sum_{j=0}^{5} \hat{\pi}_{4j} (\ln d_j) \\
1 & \sum_{j=0}^{5} \hat{\pi}_{5j} (\ln d_j)
\end{bmatrix}.$$

Notice that the elements in the second column are weighted averages of the $\ln d_j$'s, with the weights being the $\hat{\pi}_{ij}$'s. The values of these
elements for each of the five sets of misclassification probabilities are given in Table 5.23. The values in the Set #1 column are simply the values of \( \ln d_j \), \( j=0,1,...,5 \). The values in the other columns are weighted averages of the values in this first column.

**Table 5.23**

Elements in Second Column of \( \hat{\Pi}_i \) Matrices for Example 5.4.2

<table>
<thead>
<tr>
<th>Assumed Set of ( \hat{\pi}_{ij} )'s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set #1</td>
</tr>
<tr>
<td>row 1</td>
</tr>
<tr>
<td>row 2</td>
</tr>
<tr>
<td>row 3</td>
</tr>
<tr>
<td>row 4</td>
</tr>
<tr>
<td>row 5</td>
</tr>
<tr>
<td>row 6</td>
</tr>
</tbody>
</table>

As was mentioned in Section 4.4, the estimate \( \hat{\gamma} \) and its estimated standard error are not invariant to choices of \( \hat{\Pi}_i \) under this "ordinal-type" model. The values of these estimates for each set of \( \hat{\pi}_{ij} \)'s, along with the estimates of \( \tau \), are given in Table 5.24. In addition, the GOF fit statistics are not invariant. The values of these statistics are given in Table 5.25.

From Table 5.24 we can see that neither the estimates of \( \hat{\gamma} \) nor of the corresponding standard errors vary much among the five \( \hat{\pi}_{ij} \) sets. There is more variation among the estimates of \( \tau \). Notice that the
estimates \( \hat{\tau} \) increase monotonically for Sets #1 through #4 as the estimates \( \hat{\beta}_j \) for \( j=1,2,3,4,5 \) did in Table 5.20. In addition, \( \hat{\tau} \) for Set #5 is larger than for Set #4, resulting in an increase across all five sets. This is not surprising, since the estimates \( \hat{\beta}_j \), \( j=3,4,5 \), were larger for Set #5 than for Set #4, while the estimates \( \hat{\beta}_j \), \( j=1,2 \), were not much smaller for Set #5 than for Set #4.

Table 5.24

Estimates \( \hat{\tau} \) and \( \hat{\gamma} \) (and Estimated Standard Errors)

for Example 5.4.2

<table>
<thead>
<tr>
<th>Assumed Set of ( \hat{\pi}_{ij} )'s</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\tau} )</td>
<td>1.1824</td>
<td>1.2059</td>
<td>1.2527</td>
<td>1.2629</td>
<td>1.5114</td>
</tr>
<tr>
<td>( .1686 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\gamma} )</td>
<td>4.5043</td>
<td>4.5055</td>
<td>4.5048</td>
<td>4.5031</td>
<td>4.5021</td>
</tr>
<tr>
<td>( .3350 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.25 contains the standard normal variate (z) based on testing the hypothesis \( H_0: \tau=0 \) for each set of \( \hat{\pi}_{ij} \)'s. In addition, it contains 95% confidence intervals for \( \tau \). The values of the z's indicate that \( H_0: \tau=0 \) is rejected under each set of misclassification probabilities. Also notice that all these values are nearly equal. There is greater variation among the upper and lower limits of the confidence intervals, which reflects the variation among the \( \hat{\tau} \)'s and their estimated standard errors.
Table 5.25
Standardized Normal Variates (z) Based on Testing $H_0: \tau = 0$
and 95% Confidence Intervals (CI) for $\tau$ for Example 5.4.2

<table>
<thead>
<tr>
<th>Assumed Set of $\pi_{ij}$'s</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>7.0140</td>
<td>7.0222</td>
<td>7.0324</td>
</tr>
<tr>
<td>CI</td>
<td>(.8519, 1.529)</td>
<td>(.8694, 1.5424)</td>
<td>(.9036, 1.6018)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assumed Set of $\pi_{ij}$'s</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>7.0886</td>
<td>6.9139</td>
</tr>
<tr>
<td>CI</td>
<td>(.9136, 1.6022)</td>
<td>(1.0829, 1.9399)</td>
</tr>
</tbody>
</table>

From Table 5.26, we can see that, although the values of the goodness-of-fit statistics do vary, they do not vary appreciably. Each value of either $Q_p$ or $Q_{log}$ indicates the same degree of fit as the other values of that statistic.

Table 5.26
Goodness-of-Fit Statistics (and P-Values) for Example 5.4.2

<table>
<thead>
<tr>
<th>Assumed Set of $\pi_{ij}$'s</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(.8832)</td>
<td>(.8835)</td>
<td>(.8835)</td>
<td>(.8877)</td>
<td>(.8700)</td>
</tr>
<tr>
<td>$Q_{log}$</td>
<td>48.275</td>
<td>48.157</td>
<td>48.218</td>
<td>48.223</td>
<td>48.940</td>
</tr>
<tr>
<td></td>
<td>(.5825)</td>
<td>(.5873)</td>
<td>(.5848)</td>
<td>(.5846)</td>
<td>(.5559)</td>
</tr>
</tbody>
</table>

174
In conclusion, we can see that many results found when assuming no misclassification are also found when assuming various degrees of misclassification. Although $\hat{\gamma}$, its estimated standard error, and the GOF statistics are not perfectly invariant, as they were in the previous analysis, they are nearly invariant. All five $\hat{\pi}_{ij}$ sets basically agree with the choices of 4.50 and .335 for $\hat{\gamma}$ and its estimated standard error. In addition, the GOF statistic $Q_p$ indicates the same degree of goodness-of-fit for each set. The same holds for $Q_{log}$.

As was mentioned in the previous section, the estimates of $\tau$ from Sets #3 and #4 may be closest to the true estimate since there is reason to suspect that the patterns in these sets best reflect the patterns of the true misclassification probabilities. A final analysis would then include the estimates of $\tau$ from these two sets, as well as the estimate obtained when misclassification is ignored.
CHAPTER SIX

SUGGESTIONS FOR FURTHER RESEARCH

There are many issues contained in this dissertation which could be expanded upon or extended. For instance, the theory developed in Chapters 3, 4, and 5 principally focus on models which deal with situations involving misclassification of exposure alone, and no interaction. Questions similar to those answered in these chapters could be addressed in terms of other models.

Consider a model which includes interaction. It may be of interest to know whether the properties of estimators from such a model will have the same properties as those from models which don’t include interaction. For instance, the bias of \( \hat{\beta} \) and \( \hat{\gamma} \) could be determined. Also, the possible invariance of certain statistics (e.g., \( \hat{\gamma} \) and GOF statistics) case could be investigated.

In Chapter 5, we saw that models (1.16) and (1.18) work reasonably well in a variety of situations. Similar work could be done to evaluate the performance of other models, such as the model developed for misclassification of covariates and model (1.19),
developed for use with LS regression.

The basic identity upon which the models in Chapter 1 were developed, expression (1.4), was derived assuming nondifferential misclassification of exposure, a quality we believe is inherent in follow-up studies. Other types of studies, particularly case-control studies, may not induce this condition. Therefore, the development of a model which does not rely on the assumption of nondifferential misclassification could be very useful.

Further, the models developed in this dissertation allow for misclassification of exposure variables or covariates only. Models could be developed for situations in which there is misclassification of the response variable. This misclassification could occur alone, or along with misclassification of explanatory variables. The properties of the estimators and the performance of such a model could then be investigated.

Associated research can also be done to extend the ideas and results developed here to other types of statistical analyses. For example, models could be developed for survival analysis or techniques could be developed for use with analysis of survey data.

Finally, the models and results contained herein are certainly not confined to epidemiologic settings. Misclassification error occurs in a variety of research areas. If the models developed in this dissertation are not directly applicable to other types of problems or situations, adaptations to the theory can be made.
REFERENCES


