AN ABSTRACT FOR THE THESIS OF

Roger M. Sauter for the degree of Doctor of Philosophy in Statistics presented on August 18, 1989.

Title: A Method for Estimation of Generalized Linear Models When Explanatory Variables Contain Measurement Error.

Abstract approved: Redacted for Privacy

Daniel W. Schafer

This thesis considers the problem of estimating the linear parameters of generalized linear models (GLM), especially binomial and Poisson regression models, when the explanatory variable is subject to measurement error. In this situation, the dependence of the response variable on the observed explanatory variable cannot typically be modeled as a GLM; in particular, extra variability caused by measurement error cannot be accounted for using the binomial or Poisson models. One strategy is to use existing methods adapted for extra-variability. The contribution of this thesis is to introduce an estimation method which makes use of Efron's (1986) double exponential family. The proposed method involves the calculation of maximum likelihood estimates from this density when it is used as an approximation to the true density of the response variables given the observed measurements.

Efron's family of distributions offers an attractive alternative for approximating the distribution of the response variable given the

observed explanatory variable and is closely related to the measurement error in GLM methods suggested by Armstrong (1985) and Prentice (1986). Properties of the proposed method are considered when the double exponential family model is thought to be correct and when it is thought to be an approximation. Special cases and examples are given to illustrate the estimation procedure and how to apply this method. Comparisons are made with other estimation procedures for the measurement error problem, both procedurally and numerically.

A Method for Estimation of Generalized Linear Models When Explanatory Variables Contain Measurement Error

by

Roger M. Sauter

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed August 18, 1989

Commencement June 1990

APPROVED:

Redacted for Privacy

Professor of Statistics in charge of major

Redacted for Privacy

Head of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented August 18, 1989

ACKNOWLEDGEMENTS

The completion of this work was made possible with the help of others. I thank my committee members for their time and support, especially Dr. Pierce who has spent many hours in this research area blazing a trail so that others can follow.

I am greatly indebted to the help, support and encouragement of my major professor, Dr. Schafer. The time you spent with me in weekly meeting, reading the rough drafts and editing my work is very much appreciated. You did far above and beyond what could be reasonably expected of a major professor. Thank you.

A special thanks to my wife, Lori. Not only did she support and encourage me as a person during this ordeal, but she spent many hours proof reading my writing and typing. Thank you so much for helping me in so many ways, you will never know how much that meant to me.

I thank my God and Savior, Jesus Christ. Not only did He provide the strength and determination for me to attempt this work, but He also provided a peace through the times of pressure. His presence gave me the contentment to really enjoy my four years here and not just look for the day out. Self-discipline has never been my strong suit, and the Lord gave me this verse of encouragement down the stretch. "No discipline seems pleasant at the time, but painful. Latter on, however, it will produce a harvest of righteousness and peace to those who have been trained by it." Hebrews 12:11.

This research was funded by National Science Foundation grant number -- DMS 87 02111.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

A METHOD FOR ESTIMATION OF GENERALIZED LINEAR MODELS WHEN EXPLANATORY VARIABLES CONTAIN MEASUREMENT ERROR

Chapter 1

Introduction

This work concerns the estimation of parameters in generalized linear models when the explanatory variable contains measurement error. The naive method, which ignores explanatory variable measurement error, leads to inconsistent estimators of the parameters (Stefanski and Carroll, 1985). The proposed method approximates the likelihood based on the distribution of the responses given the measurements by using Efron's (1986) double exponential family distribution and obtains estimates that maximize this approximation to the likelihood. The proposed method is closely related to estimators presented by Armstrong (1985) and Prentice (1986). Section 1 of this chapter discusses the basic problem of measurement error in the explanatory variable. Examples are introduced in section 2. The last section of this chapter outlines the remaining elements of this thesis.

1.1 Measurement Error in the Explanatory Variable

Generalized linear models are used to model discrete data in the form of binomial and Poisson regression. These models, like their continuous counterparts in ordinary regression, assume that the explanatory variable is known or observed exactly. Many researchers have examined the problem of estimating regression parameters in the presence of explanatory variable measurement error. Some methods of estimation for ordinary regression are discussed in chapter 2. Chapter 3 considers some methods of estimation for this problem in a generalized linear model setting. Although the thrust of this thesis is to propose a new method of estimation, it includes comparisons of the proposed method to existing methods. Simulated comparisons focus on the performance of the methods on an example introduced in section 1.2.1 below. Data analysis is conducted on both of the examples introduced in the next section.

1.2 Introduction to the Examples

1.2.1 Chromosome Aberration Data Set

The first data example is of 649 survivors of the atomic bomb dropped on Hiroshima in 1945 (Otake and Prentice, 1984). The response variable for each individual is the proportion of cells, out of 100 examined, that had chromosome aberrations. The explanatory variable of interest is the total exposure to radiation for an individual. The measurement of total radiation exposure for an individual involved physical calculations based on his or her distance from the blast and the shielding configuration. The data for the calculations was gathered through interviews with 649 survivors. A goal in the analysis of this data is to estimate the regression of the proportion of chromosome aberrations, denoted by p, on the true dose of radiation exposure. It is desired to model p as a linear function of the total radiation exposure:

$$p_i = \beta_0 + \beta_1 \text{exposure}_i.$$

Models that are quadratic in exposure are also of some interest but will not be examined here. It is known that substantial measurement error is present in the observed values of radiation. Figure 1 on page 4 contains a scatter plot of the data. The data was grouped into seven categories, where the zero radiation exposure category is thought of as a control group. A small random uniform number was added to each grouped average exposure to display the concentrations of points in each category. The raw grouped data is listed in Appendix A.

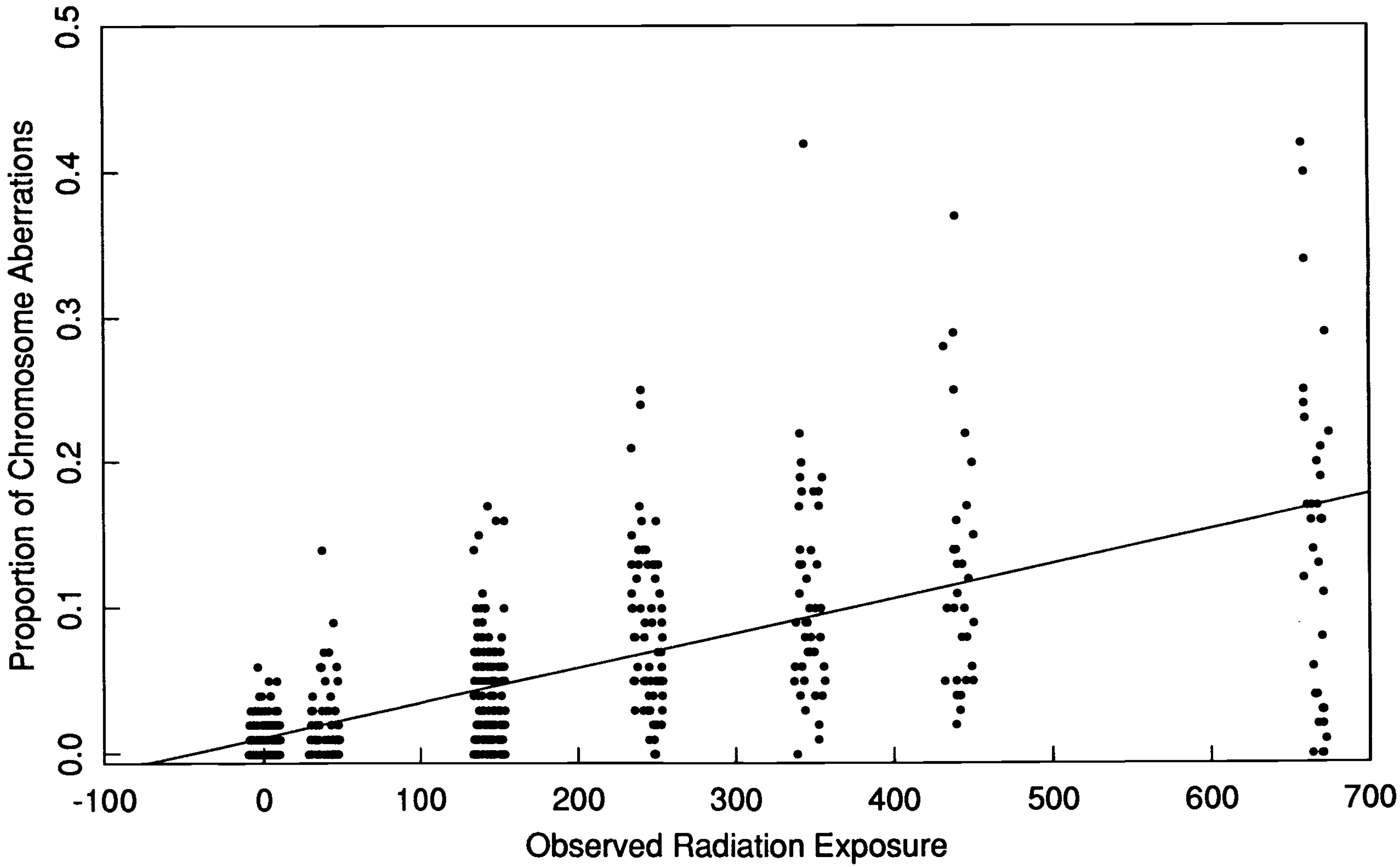
1.2.2 Coronary Heart Disease Data Set

This example is taken from a study by Iso et al. (1989) to investigate the association of serum cholesterol and age-adjusted rates of death from coronary heart disease for 350,977 men between the ages of 35 and 57 years old. Although a more formal analysis would take age into account explicitly (as well as smoking habits), it is convenient to consider the simple model with age-adjusted rate of coronary heart disease as a response variable and serum cholesterol as an explanatory variable which is measured with error.

As a starting point, suppose that the probability of coronary heart disease, p, is related to cholesterol by

$$\log(p_i) = \beta_0 + \beta_1 \text{cholesterol}_i.$$

Figure 1. Plot of Chromosome Aberration Data.



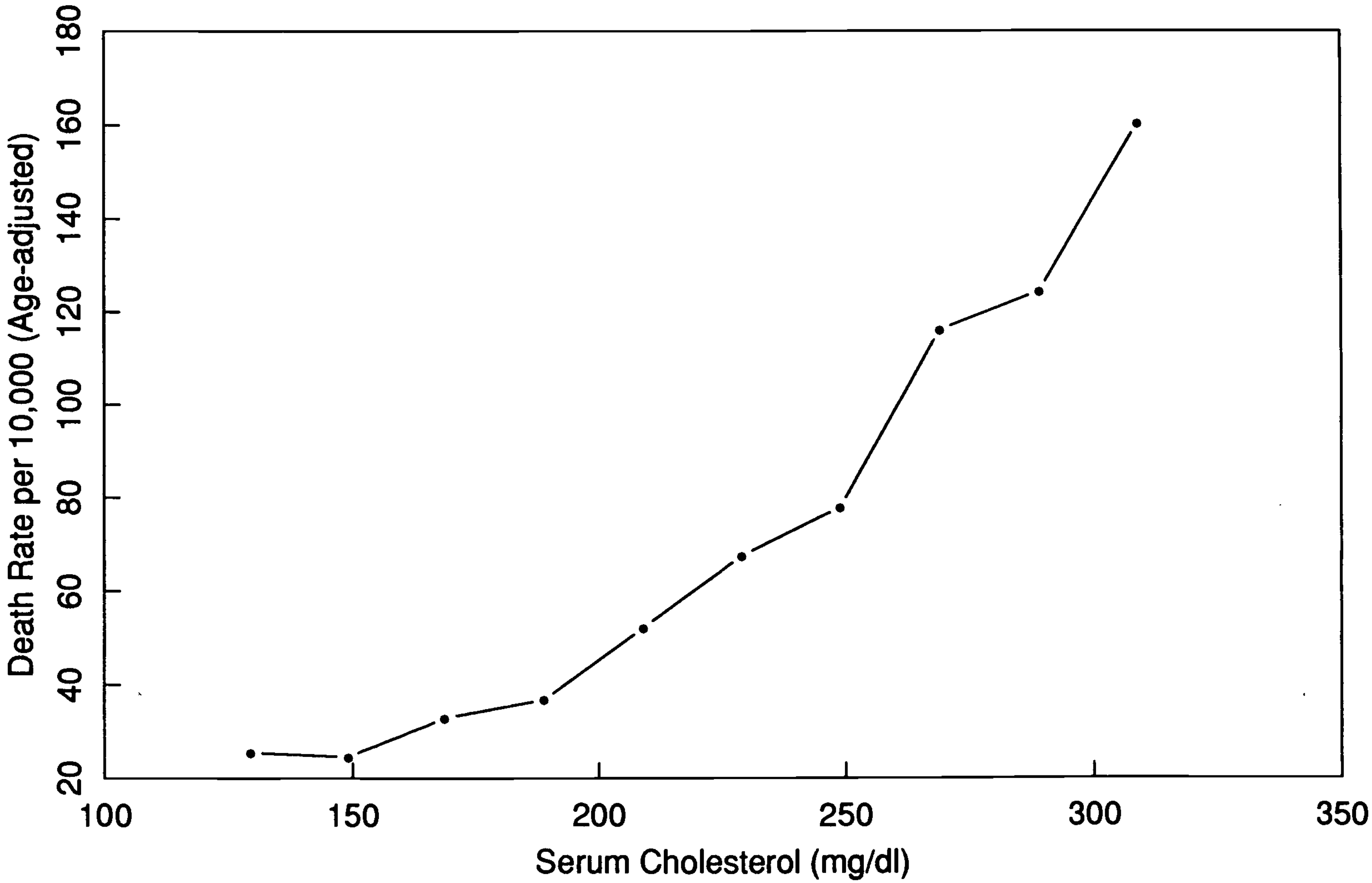
For a large number of individuals, m_i, with approximately the same cholesterol level, the number of individuals with coronary heart disease may be modeled as Poisson with mean $m_i p_i$. The age-adjusted rate is modeled as having mean p_i and variance p_i/m_i. It is observed that the measurement error in cholesterol levels can have a substantial effect on the estimate of β. This data has been grouped into 10 cholesterol levels. A graph of the data is presented in Figure 2 and the data is listed in Appendix A. The double exponential family is used to model the data to illustrate the mechanics of using the method.

1.3 Outline of Thesis

This thesis proposes a method for estimating the parameters in generalized linear models when the observed explanatory variable contains measurement error and identifies the properties of the proposed estimator. Some asymptotic results for an approximation to the proposed method are presented and small sample properties are compared in several situations of interest, using simulation. Relevant comparisons of the proposed method to existing methods, particularly Armstrong's (1985) method and Prentice's (1986) method, include a simulation comparison to judge relative performance on a simulated model which is comparable to the chromosome aberration data.

A literature search is presented in chapters 2 through 4. Chapter 2 cites work conducted to address the measurement error problem in ordinary regression. Methods for estimation in the presence of measurement error for generalized linear models are

Figure 2. Plot of Coronary Heart Disease Data.



discussed in Chapter 3. In chapter 4, the concept of overdispersion is introduced and methods developed in this area are presented, focusing on the double exponential family. Chapter 5 introduces the proposed estimation procedure using the double exponential family. Chapter 6 contains the analysis of the data sets. A numerical comparison of methods is presented in chapter 7. The last chapter summarizes the conclusions and discusses further topics of interest.

Chapter 2

Measurement Error in Ordinary Regression

2.1 Introduction

A large body of literature exists for the problem of measurement error in explanatory variables. This chapter offers a brief overview of some issues regarding the errors-in-variables problem without providing exhaustive coverage. Fuller's book (1987) contains a modern review and list of references for work on errors-in-variables.

In order to discuss some relevant issues concerning measurement error in simple linear regression, it is convenient to label the following set of assumptions as MODEL 1:

1) (y_1, z_1), $(y_2, z_2), \ldots, (y_n, z_n)$ are independent pairs of random variables

2) $y_i = \beta_0 + \beta_1 x_i + e_i$, for $i=1,\ldots,n$,

3) $z_i = x_i + d_i$, and

4) e_i and d_i are independent random variables with means 0 and variances σ_e^2 and σ_d^2, respectively

where y_i is the response variable,

x_i is the true but unknown explanatory variable,

z_i is the observed explanatory variable, and

d_i is the random measurement error.

Some additional assumptions are used to illustrate specific points. The following assumptions, in addition to MODEL 1, are referred to as

MODEL 1A:

$$(e_i, d_i, x_i)' \sim N[(0,0,\mu_x)', \ \mathrm{diag}(\sigma_e^2, \sigma_d^2, \sigma_x^2)].$$

2.2 The Problem with Ignoring Measurement Error

There is one situation for which the presence of measurement error in the explanatory variable can be ignored. When the purpose is to estimate an equation for predicting future y's from future imprecisely measured z's with the same measurement error distribution, the presence of measurement error in the explanatory variable can be ignored. In this situation, the relationship between the y's and z's is the relationship of interest.

However, measurement error in the explanatory variable cannot be ignored when it is desired to estimate the coefficients in the regression of Y on the true explanatory variable. The following example is used to illustrate a problem that arises when measurement error is ignored. Consider MODEL 1A; the vector $(y_i, z_i)'$ is bivariate normal with the following properties:

$$E\{(y_i, z_i)\} = (\mu_y, \mu_z) = (\beta_0 + \beta_1\mu_x, \mu_x), \text{ and}$$

covariance matrix

$$\begin{bmatrix} \sigma_y^2 & \sigma_{zy} \\ \sigma_{zy} & \sigma_z^2 \end{bmatrix} = \begin{bmatrix} \beta_1\sigma_x^2 + \sigma_e^2 & \beta_1\sigma_x^2 \\ \beta_1\sigma_x^2 & \sigma_x^2 + \sigma_d^2 \end{bmatrix}.$$

The naive estimator for β_1, denoted by $\hat{\beta}_1$, is the least squares estimate of slope in the simple regression of y_i on z_i. One problem caused by ignoring the measurement error is that this estimator is biased, i.e.

$$E[\hat{\beta}_1] = \beta_1(\sigma_x^2/(\sigma_x^2+\sigma_d^2)) = \beta_1(\sigma_x^2/\sigma_z^2).$$

The ratio $R = \sigma_x^2/\sigma_z^2$ is referred to as the reliability ratio. The least squares solution, $\hat{\beta}_1$, is biased towards zero (i.e. the measurement error attenuates the regression coefficient). As the variance of the measurement error gets larger relative to the variance of the x_i, the usual least squares estimator becomes increasingly biased.

2.3 <u>The Need for Additional Information</u>

Consider again MODEL 1A in which e_i, d_i, and x_i are all normally distributed. In this situation, jointly sufficient statistics for the distribution of y and z are

$$m_y, m_z, m_{yy}, m_{yz}, m_{zz}, \text{ where } m_y = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ and } m_{yz} = \frac{1}{n}\sum (y_i-\bar{y})(z_i-\bar{z}).$$

There are five sufficient statistics for six parameters $(\mu_x, \sigma_x^2, \sigma_e^2, \sigma^2, \beta_0, \beta_1)$ and therefore the parameters are not identifiable. More information is needed to estimate all the parameters.

Although there are models other than the normal for which the parameters are identifiable, the basic problem remains that little can be said about the regression coefficients without some additional information about one of the parameters. The extra information may be internal. For example, if there are replicate values of z at several x's, the parameters are identifiable. When there are no such replicates, estimation of the regression coefficients depends on the availability of outside estimates (or knowledge) of the parameters, typically σ_d^2 or R.

Missing information can be provided in a number of ways. When the reliability ratio, R, is known, the parameter space is reduced to five dimensions and the maximum likelihood estimators may be obtained by equating the sufficient statistics to their expectations (Kendall and Stuart, Ch. 29) and the unbiased estimate of β_1 is $\tilde{\beta}_1 = R^{-1}\hat{\beta}_1$.

If the measurement error variance, σ_d^2, is known, the dimension of the parameter space is five and the maximum likelihood estimators are given by

$$\hat{\beta}_1 = (m_z - \sigma_h^2)^{-1} m_{zy},$$

$$(\hat{\sigma}_x, \hat{\sigma}_e) = (m_z - \sigma_h^2, m_y - \hat{\beta}_1 m_{zy}),$$

$$(\hat{\mu}_x, \hat{\beta}_0) = (\bar{z}, \bar{y} - \hat{\beta}_1 \bar{z}).$$

Another way that needed information can be provided is through an instrumental variable, w_i, which is independent of e_i and d_i, but correlated with x_i. Suppose, in addition to MODEL 1A, that w_i is a random variable with mean μ_w and variance σ_w^2 which is independent of e_i and d_i and that $\text{cov}(w_i, x_i) = \sigma_{xw} \neq 0$. This leads to the following,

$$\begin{bmatrix} x_t \\ e_t \\ d_t \\ w_t \end{bmatrix} \sim NI\left(\begin{bmatrix} \mu_x \\ 0 \\ 0 \\ \mu_w \end{bmatrix}, \begin{bmatrix} \sigma_x & 0 & 0 & \sigma_{xw} \\ 0 & \sigma_e & 0 & 0 \\ 0 & 0 & \sigma_d & 0 \\ \sigma_{xw} & 0 & 0 & \sigma_w \end{bmatrix} \right)$$

where $\mu_x = \alpha_0 + \alpha_1 \mu_w$ and it is assumed that $\sigma_{xw} = \alpha_1 \sigma_w^2$. Under these assumptions, the observed vector (y_i, z_i, w_i) is normal with mean vector

$$(\beta_0 + \beta_1 \alpha_0 + \alpha_1 \mu_w,\ \alpha_0 + \alpha_1 \mu_w,\ \mu_w),$$

covariance matrix

$$\begin{bmatrix} \sigma_e^2 + \beta_1^2 \sigma_x^2 & \beta_1 \sigma_x^2 & \beta_1 \alpha_1 \sigma_w^2 \\ \beta_1 \sigma_x^2 & \sigma_x^2 + \sigma_d^2 & \alpha_1 \sigma_w^2 \\ \beta_1 \alpha_1 \sigma_w^2 & \alpha_1 \sigma_w^2 & \sigma_w^2 \end{bmatrix},$$

and nine parameters with a set of nine minimal sufficient statistics. Thus, the parameters are identifiable and maximum likelihood estimation can be used.

This section has focused on MODEL 1A. All forms of additional information discussed here can be used in the general setting of MODEL 1. Fuller (1987) lists the asymptotic properties of these in a general setting.

2.4 Structural and Functional Models

The structural model refers to MODEL 1 with the assumption that the x_i's are independent and identically distributed. The functional model assumes that the x_i's are fixed constants.

In ordinary regression, without measurement error, the typical assumptions for the model are

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where β_0 and β_1 are parameters,

x_i are known constants, and

e_i are random variables with mean 0 and variance σ_e^2.

With these assumptions, the least squares estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, are the best linear unbiased estimators. These same properties hold when the x_i's are independent and identically distributed random variables with these assumptions:

1) the conditional distribution of y_i given x_i has mean, $\beta_0+\beta_1 x_i$, and variance, σ_e^2, and
2) the marginal distribution of x_i does not involve β_0, β_1, σ_e^2.

So, except for the rare possibility of (2) above, there is no need in ordinary regression to be concerned with whether or not the x_i's are realizations of a random variable or fixed contstants.

For measurement error models, the distinction between functional and structural models is more important because, for the functional model, $x_1,\ldots,x_n$ are unknown constants, and therefore parameters. The structural model is one way to eliminate the n incidental parameters by assuming they come from a common probability distribution depending on only a few parameters. There are some important differences between these two models. The structural model is usually easier to incorporate into standard statistical methods. In theory, it is possible to find the maximum likelihood estimates of the parameter and the asymptotic distribution of these estimators. The functional model's drawback is that the number of incidental parameters increases with an increase in the sample size. Thus, maximum likelihood estimates are not appropriate or not useful for the functional model (Kendall and Stuart, Chapter 29). Method of moments estimators are available for the functional model but it is difficult to find the asymptotic distribution of the estimators without using an asymptotic framework where the measurement error variance approaches zero. One drawback to the structural model is that it is only appropriate when the explanatory variables are independent and identically distributed. This is a much stronger condition than is generally required in regression analysis. In addition, the distributions for the independent and identically distributed x's and d's, for which maximum likelihood estimators can be found, are very limited, while the functional model may be used no matter how the x_i are selected.

2.5 Nonlinear Models

The situation to be considered in this section is the nonlinear model. This model is as follows:

$$y_i = g(x_i, \beta) + e_i, \quad z_i = x_i + d_i, \text{ and}$$

$$(e_i, d_i) \sim NI[0, \text{diag}(\sigma_e^2, \sigma^2)],$$

where e_i, d_i and x_i are independent. Fuller (1987) uses a Taylor expansion of $g(x_i,\beta)$ about $g[E(x_i|z_i),\beta]$ to arrive at the following approximation for the mean of the distribution of y_i given z_i, when x_i is also assumed to be normal:

$$E(y_i|z_i) \doteq g[E(x_i|z_i),\beta] + \tfrac{1}{2}\mathrm{tr}\{g''[E(x_i|z_i),\beta]\cdot \mathrm{Var}(x_i|z_i)\}$$

He uses this as his new model, i.e.

$$y_i = g[E(x_i|z_i),\beta] + (0.5)\mathrm{tr}\{g''[E(x_i|z_i),\beta]\cdot \mathrm{Var}(x_i|z_i)\} + e_i^*.$$

and suggests using an iterative weighted nonlinear least squares method on this model to estimate the parameters. The first step is to use ordinary nonlinear least squares to arrive at initial estimates of parameters. The second step is to determine the weights from these estimates for the third step of weighted nonlinear least squares. Once the new estimates are calculated, the second step is repeated to update the weights, so that step three can be performed again. Steps two and three are repeated until convergence occurs. For more details see Fuller (1987).

2.6 The Berkson Model

The following model, due to Berkson (1950), is different than the classical measurement error model but is relevant in some later discussions. Suppose that y_i and z_i are observed, where

$$y_i = \beta_0 + \beta_1 x_i + e_i \text{ and}$$

$$x_i = z_i + d_i,$$

where the n pairs (e_i, d_i) are independent with means (0,0) and variances (σ_e^2, σ_d^2), and where d_i is independent of z_i. The latter point is the distinguishing characteristic of the Berkson model and implies that $E(x_i|z_i) = z_i$. In this situation, the z_i's are controlled by the experimenter and can be thought of as fixed. The model can be written as,

$$y_i = \beta_0 + \beta_1 x_i + e_i = \beta_0 + \beta_1 (z_i + d_i) + e_i = \beta_0 + \beta_1 z_i + v_i$$

where $v_i = e_i + \beta_1 d_i$ and z_i and v_i are independent. This implies that

$$E(y_i) = \beta_0 + \beta_1 z_i.$$

Berkson realized that the ordinary least squares estimates can be used when z_i is fixed. The least squares estimates are

$$\hat{\beta}_1 = \left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]^{-1}\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y}) \text{ and}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{z}.$$

Because the z_i are fixed, it is straightforward to show that these estimates are unbiased, as follows,

$$E(\hat{\beta}_1) = E\left\{\left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]^{-1}\sum_{i=1}^{n}(z_i-\bar{z})(y_i-\bar{y})\right\}$$

$$= \left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]^{-1}\sum_{i=1}^{n}\{(z_i-\bar{z})[E(y_i)-E(\bar{y})]\}$$

$$\left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]^{-1}\sum_{i=1}^{n}\{(z_i-\bar{z})[(\beta_0+\beta_1 z_i)-(\beta_0+\beta_1\bar{z})]\}$$

$$\left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]^{-1}\sum_{i=1}^{n}(z_i-\bar{z})^2\beta_1 = \beta_1$$

and similarly for $\hat{\beta}_0$. Thus, if $(e_i,d_i)' \sim NI(0,\mathrm{diag}(\sigma_e^2,\sigma_d^2))$, then the usual inference with β_0 and β_1 can be carried out. Estimation is unbiased only for the linear model. It is possible to obtain biased results if the same measurement error holds for several replicates in replicated experiments. Also, if the mean of y_i is not linear in x_i, least squares estimation can give biased estimators.

Chapter 3

Measurement Error in Generalized Linear Models

This chapter is an overview of the previous literature on the problem of measurement error in explanatory variables for generalized linear models (GLMs). The first section introduces the notation that is used for GLMs in the rest of this work. The following two sections review specific estimators for functional and structural assumptions. The fourth section introduces the role of overdispersion in the measurement error problem.

3.1 Notation

Suppose y_i is a response variable whose distribution comes from the one-parameter exponential family,

$$f_y(y_i) = \exp\{[y_i\eta_i+b(\eta_i)](m_i/\phi)+c(y_i,\phi)\},$$

where $b'(\eta_i)=\mu_i$ is the mean of y_i, ϕ is the scale or dispersion parameter, and m_i is a known constant. The variance function is defined as $V(\mu_i) = b''(\eta_i)$ and implies that

$$\mathrm{Var}(y_i) = (\phi/m_i)V(\mu_i).$$

The mean of y_i is assumed to be related to an explanatory variable x_i through a nonlinear model of the following form:

$$E(y_i|x_i) = \mu_i= g^{-1}(\beta_0+\beta_1x_i), \tag{3.1}$$

where β is the 2 x 1 vector of parameters and $g(\cdot)$ is referred to as the link function. It is also assumed that the y_i's are independent.

Suppose that z_i is a measurement of x_i and that $f(y_i|x_i) = f(y_i|x_i,z_i)$. This condition implies that z_i is useful in predicting y_i only through x_i. Thus, when x_i is known, z_i provides no new information about y_i. It is straightforward to find maximum likelihood estimates for β when x_i is known (see McCullagh and Nelder (1983) for details). Because x_i is observed only through z_i, it is necessary to consider models to represent this relationship. The following sections review some specific examples of what has been done to deal with this problem.

3.2 Functional Model Estimators

This section introduces some functional model estimators that were suggested by Stefanski (1985) and Stefanski and Carroll (1985). The inherent problem of having a functional model is that the fixed x_i's are unknown parameters. The problem does not resolve itself with an increase in sample size because the number of parameters increases correspondingly.

3.2.1 Stefanski's (1985) Approach

This approach for dealing with measurement error is to modify a naive estimator, $\hat{\beta}$. Stefanski assumes that β can be estimated by an M-estimator (including maximum likelihood) which is consistent when the true x_i are known. This M-estimator, $\tilde{\beta}$, is the solution to

$$\sum_{i=1}^{n} \psi(y_i, x_i, \tilde{\beta}) = 0$$

where, for maximum likelihood estimation, ψ would be a log-likelihood score function. It is assumed that the y_i's are independent and that

$$E[\psi(y_i, x_i, \beta)] = 0, \quad \text{for } i=1,\ldots,n.$$

A naive estimator, $\hat{\beta}$, of β can be found by ignoring measurement error, as the solution to

$$\sum_{i=1}^{n} \psi(y_i, z_i, \hat{\beta}) = 0.$$

The estimator, $\hat{\beta}$, converges in probability to β^* which is the solution to

$$\lim n^{-1} \sum_{i=1}^{n} E\{\psi(y_i, z_i, \beta^*)\} = 0.$$

Generally $\beta^* \neq \beta$, therefore $\hat{\beta}$ is asymptotically biased. Stefanski's approach is to approximate the asymptotic limit of $\hat{\beta}$, in an asymptotic setting where the measurement error is small, and subtract the estimate from $\hat{\beta}$ in order to arrive at an estimator that is less biased. This approach is illustrated in the following example. For the simple model

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad \text{and} \quad z_i = x_i + d_i,$$

where d_i is independent of x_i and e_i, and d_i has mean 0 and variance

σ_d^2, the limit of the naive estimator, $\hat{\beta}$, is

$$\beta^* = \{\lim n^{-1}\sum_{i=1}^{n}[E(x_i x_i')] + E(d_i d_i')\}^{-1}\{\lim n^{-1}\sum_{i=1}^{n}[E(x_i x_i')]\}\beta,$$

where $x_i = \begin{bmatrix}1\\ x_i\end{bmatrix}$ and $d_i = \begin{bmatrix}0\\ d_i\end{bmatrix}$. The corrected estimator suggested by Stefanski is thus,

$$\hat{\beta}^* = \left\{I + \begin{bmatrix}1 & m_z\\ m_z & m_{zz}\end{bmatrix}^{-1}\begin{bmatrix}0 & 0\\ 0 & \sigma_d^2\end{bmatrix}\right\}\hat{\beta}.$$

He shows that this estimator has less bias than the naive estimator, that it is consistent, and that, under regularity conditions, it is asymptotically normal. The expressions for other generalized linear models are lengthy and not reproduced here.

3.2.2 Stefanski and Carroll's (1985) Approach

Stefanski and Carroll (1985) proposed four functional model estimators. All of these approaches deal with logistic regression , such that x_i are fixed predictors and y_i are Bernoulli variates with

$$\Pr\{y_i=1|x_i\} = F(x_i'\beta) = (1 + \exp(-x_i'\beta))^{-1},\ i = 1, 2,\ldots,n, \quad (3.2)$$

where $x_i' = [1\ x_i]$. The approach starts with the usual logistic regression estimator of β, which is asymptotically biased, and then introduces modifications to this estimator. All modifications involve estimating both vectors β and x because the functional model is used. The initial estimate is obtained by regressing y_i on z_i. This estimator is called $\hat{\beta}$ and satisfies

$$\sum_{i=1}^{n}[y_i - F(c_i'\hat{\beta})]c_i = 0,$$

when $c_i = z_i$. The variable c_i is the place holder for the explanatory variable to be used and z_i is the observed explanatory variable. If the x_i were available, $c_i = x_i$ would lead to the usual maximum likelihood estimates. Stefanski and Carroll show that $\hat{\beta}$ is inconsistent for β. Using $\hat{\beta}$ as the starting point, three alternative approaches are considered by using different choices of c_i.

Each of these modified estimators can be thought of as a type of two-stage estimator obtained by doing logistic regression with $\hat{x}_i$ replacing z_i and then maximizing the log-likelihood based on (3.2) with respect to β. The first modification is distribution free in the sense that only moment assumptions are made about the measurement error. In this approach adjustment for the bias in $\hat{\beta}$ is made by maximizing (3.2) using $c_i = \hat{x}_i$, where

$$\hat{x}_i = z_i + \hat{\sigma}^2(I - \hat{\sigma}^2\hat{B}_n)^{-1}\hat{B}_n z_i,$$

where $\hat{B}_n$ is a correction for asymptotic bias as $\min(n, \sigma^{-1}) \longrightarrow \infty$ which depends only on the observed data (See Stefanski and Carroll, 1985, for details).

The other modifications suggested by Stefanski and Carroll (1985) add an additional assumption of normality of the d_i to the model (3.2). The second modification uses the log-likelihood for estimating β and x, based on these assumptions, to estimate $c_i = \hat{x}_i$, defined as

$$\hat{x}_i = z_i + [y_i - F(z_i'\hat{\beta})]\hat{\sigma}^2\hat{\beta}$$

In the last modification considered, Stefanski and Carroll take σ and β to be known and find a sufficient statistic for estimating x_i. The particular sufficient statistic they choose has the property of making the estimate for β unbiased. Here $c_i = \hat{x}_i$

$$\hat{x}_i = z_i + \hat{\sigma}^2(y_i - \tfrac{1}{2})\hat{\beta}.$$

The Stefanski and Carroll (1985) methods extend to models with additional explanatory variables.

3.3 Structural Model Estimators

This section considers structural models in GLM. Five particular estimators are introduced. The last two, covered in sections 3.3.4 and 3.3.5, are very close in spirit to the method proposed in chapter 5 and are discussed in greater detail. They are also compared in more depth in later chapters both numerically and procedurally.

3.3.1 Schafer's (1987) Approach

This method involves treating the true explanatory variables, x_i, as missing data. The idea is to use the EM algorithm to obtain the MLE for β where the likelihood is based on the joint distribution of y and z. It is assumed that z_i is distributed normal(x_i, σ_z^2) and that x_i is distributed normal(μ_x, σ_x^2). The general form for this procedure of estimation is accomplished by using the following two steps at the t^{th} iteration:

E-step: Compute $Q(\beta|\beta^{(t)}) = E[\log f(y,z,x,S_m|y,z,S_m,\beta_{(t)})]$, which requires computation of approximate first and second conditional moments of x_i given z_i and where S_m is an unbiased estimate of the measurement error sample covariance matrix.

M-step: Choose $\beta^{(t+1)}$ to maximize $Q(\beta|\beta^{(t)})$, where regression parameters are updated by iteratively weighted least squares.

Approximations are suggested to simplify these calculations.

Schafer concluded that this approximation to the maximum likelihood estimator based on normality did well in almost all conditions that he examined and may be useful more generally. Whittemore and Keller (1988) reported that error for this estimator can be reduced to the order of $o(\delta)$, where δ is an arbitrarily small constant.

3.3.2 Whittemore and Keller's (1988) Method

A method of adjusting the naive estimator of an approximate likelihood function to improve estimation in the presence of measurement error was proposed by Whittemore and Keller (1988). A likelihood score vector $\ell(y_i|z_i;\beta,\delta)$ is proposed where assumptions only involve the mean and variance of the measurement error conditional on the observed explanatory variable, z_i. These assumptions are valid when the moments are small, say of order δ, where δ is a small parameter representing the size of the measurement error variance, and when the third moments are of order $o(\delta)$.

Whittemore and Keller define the ideal maximum likelihood estimate to be $\hat{\beta}(\delta)$, the solution to the following equations,

$$L(\hat{\beta},\delta) = \sum_{i=1}^{n} \ell(y_i|z_i;\hat{\beta},\delta) = 0. \tag{3.3}$$

When $\delta = 0$, the solution $\hat{\beta}(0)$ is the naive estimator which ignores measurement error. This is inconsistent and has bias of order $O(\delta)$. In order to help correct this bias, Whittemore and Keller propose an estimator that depends on the first two terms of a Taylor expansion for $\hat{\beta}(\delta)$ about 0. The computations involve taking the derivative of (3.3) with respect to δ, evaluating at δ=0, and solving for $\hat{\beta}(0)$, this leads to the following estimator,

$$\beta^{*} = \hat{\beta}(0) + \delta\{-L_{\beta}[\hat{\beta}(0),)]\}^{-1}L_{\delta}[\hat{\beta}(0),0].$$

For details of this procedure, examples, and conditions for β^{*} to be consistent, see Whittemore and Keller (1988).

3.3.3 Carroll, Spiegelman, Lan, Bailey, and Abbott's (1984) Approach

For binary regression, Carroll et al. (1984) start with,

$$p_i = pr(y_i = 1|x_i) = G(\beta_0 + \beta_1 x_i)$$

where $G(\cdot)$ is a known distribution function like $G(a) = (1 + e^{-a})^{-1}$ for logistic regression or $G(a) = \Phi(a)$ for Probit regression. Here $\Phi(\cdot)$ is a standard normal distribution function. The observed explanatory variable is $z_i = x_i + d_i$, where d_i is the measurement error and it is assumed to be independently and normally distributed with mean 0 and variance σ_d^2.

Carroll et al. eliminate the nuisance parameters x_i by assuming that x_i are independently and normally distributed with mean μ_x and variance σ_x^2. Initially, the parameters μ_x, σ_x^2 and σ_d^2 are assumed known. Carroll et al. condition on the observed values z_i and use the likelihood of y_i conditioned on z_i for estimation. For logistic regression, this involves numerical integration but for Probit regression it can be evaluated explicitly. In practice, the parameters μ_i and σ_x^2 are estimated in the first stage of the analysis from the z_i's and it is assumed that σ_d^2 has a specific value.

3.3.4 Armstrong's (1985) Approach

Armstrong's approach is to force the induced model for y_i given z_i into the same distributional form as $f(y_i|x_i)$ but with mean and variance to match $E(y_i|z_i)$ and $Var(y_i|z_i)$. These are given by

$$E(y_i|z_i) = E(E(y_i|x_i)|z_i) = E(g^{-1}(x_i'\beta)|z) = \mu_i^*,$$

$$\begin{aligned} \text{and} \quad Var(y_i|z_i) &= Var(E(y_i|x_i)|z_i) + E(Var(y_i|x_i)|z_i) \\ &= Var(g^{-1}(x_i'\beta)|z_i) + E(\frac{\phi}{w_i}V(\mu_i)|z_i) \\ &= Var(\mu_i|z_i) + \frac{\phi}{w_i}E(V(\mu_i)|z_i). \end{aligned} \tag{3.4}$$

In order to fit this into the original setting, Armstrong approximates

$$E(g^{-1}(x_i'\beta)|z_i) = \mu_i^* \approx g^{*-1}(z_i'\beta),$$

where g^{*-1} is the approximating function that relates μ_i^* to $z_i'\beta$ not necessarily the same as g^{-1}. Next, he equates the desired variance, $\mathrm{Var}(y_i|z_i) = \frac{\phi}{w_i^*}V(\mu_i^*)$, to (3.4), which may be accomplished by taking as a prior weight

$$w_i^* = \frac{\phi V(\mu_i^*)}{(\phi/w_i)E(V(\mu_i)|z_i) + \mathrm{Var}(\mu_i|z_i)},$$

where μ_i, w_i, ϕ, $V(\cdot)$ are the mean, weight, scale parameter, and variance function of the original GLM for y_i given x_i. This gives the following set-up:

$$E(y_i|z_i) = \mu_i^* = g^{*-1}(z_i'\beta), \quad \text{and}$$

$$\mathrm{Var}(Y_i|z_i) = \frac{\phi}{w_i^*}V(\mu_i^*).$$

Once w_i^* is estimated, Armstrong assumes w_i^* is known and the estimation of β is accomplished by maximum quasi-likelihood estimation using GLIM (Baker and Nelder, 1978).

3.3.5 Prentice's (1986) Approach

This method, which is appropriate for binomial response variables, is similar to Armstrong's except that the induced model for y_i given z_i is approximated by a beta-binomial distribution rather than a generalized linear model. The mean-variance relationship for a beta-binomial random variable Y is given by:

$$E(Y) = mp \quad \text{and}$$

$$\begin{aligned} Var(Y) &= mp(1-p)[1+(m-1)\rho] \\ &= \frac{mp(1-p)}{\theta}, \end{aligned}$$

where $\theta^{-1} = 1+(m-1)\rho$. This relationship results when $Y = \sum_{j=1}^{m} U_j$ where the U_j are correlated Bernoulli random variables with correlation, ρ. It is also the mean-variance relationship one would find if the U_j were independent Bernoulli trials with the probabilities p_j being random variables having a beta distribution (see Williams, 1982). For many purposes, the extra parameter θ offers a convenient way of incorporating extra-binomial variation and maximum likelihood can be used to estimate the unknown parameters.

Prentice considers a special regression problem where measurement error in the explanatory variables is the sole source of extra-binomial variation. In this case, regression models are induced for both $p(z_i)$ and $\theta(z_i)$. He embeds the resulting expressions for $E(y_i|z_i)$ and $Var(y_i|z_i)$ into a beta-binomial model so that maximum likelihood can be used for estimation. The likelihood function, $L(p,\theta)$, for a beta-binomial, is used to estimate the parameters of interest, β, which are the coefficients from the induced regression model. The score statistic and information are obtained through

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial p}\frac{\partial p}{\partial \beta} + \frac{\partial L}{\partial \theta}\frac{\partial \theta}{\partial \beta}$$

and a similar expression for $-\frac{\partial^2 L}{\partial \beta \partial \beta'}$. These expressions, along with the first and second derivatives of $p(z_i)$ and $\theta(z_i)$ with respect to β, make it possible to carry out standard asymptotic likelihood inference on β (note, however, that the beta-binomial model itself is only a useful approximation to the distribution of y_i given z_i in this situation).

3.4 The Role of Overdispersion

One facet of this problem, implicit in the two previous sections, is that the variation of y_i given z_i is greater than the variation of y_i given x_i. The problem is illustrated using the model given in (3.1) where $g(\cdot)$ is the identity function and the distribution of y_i given x_i is binomial$(m,p_i)/m$. These assumptions lead to the following mean-variance relationship for the distribution of y_i given x_i:

$$E(y_i|x_i) = p_i = \beta_0 + \beta_1 x_i, \text{ and}$$

$$\mathrm{Var}(y_i|x_i) = \frac{p_i(1-p_i)}{m} \qquad \text{for } i=1,\ldots,n.$$

The mean-variance relationship needed for estimation is the conditional distribution of y_i given z_i, as follows:

$$E(y_i|z_i) = p_i^* = \beta_0 + \beta_1\cdot E(x_i|z_i) \tag{1}$$

and

$$\mathrm{Var}(y_i|z_i) = \frac{1}{m}\cdot p_i^*(1 - p_i^*) + (1-\frac{1}{m})\cdot\beta_1^2\cdot\mathrm{Var}(x_i|z_i),$$

$$= \frac{p_i^*\cdot(1 - p_i^*)}{m\cdot\theta_i}, \tag{2}$$

$$\text{where} \quad \theta_i^{-1} = 1 + \beta_1^2 \frac{\operatorname{var}(x_i \mid z_i)}{p_i^*(1 - p_i^*)}.$$

Written in this form, the distribution of y_i given z_i looks similar to the distribution of y_i given x_i. The main difference between the two is the replacement of x_i by $E(x_i|z_i)$ in the mean and the addition of the overdispersion factor, θ_i. One consequence of using the measured explanatory variable in place of the true values is that the variability of y_i conditional on the measured variables is greater than what is expected under the binomial model. If the density function for the conditional distribution of y_i given z_i were available, estimates of the unknown parameters could be obtained by maximum likelihood. For the model given, however, the density cannot be expressed in closed form. The idea is to use a density which approximates $f(y_i|z_i)$ and has the correct mean and variance as given above. The quasi-likelihood method of section 3.3.4 and the beta-binomial structure of section 3.3.5 are two methods to estimate parameters in the presence of overdispersion. An additional method, based on the double exponential family, is introduced in the next chapter.

Chapter 4

Estimation in Overdispersed Generalized Linear Models and an Introduction to the Double Exponential Family

While generalized linear models have proven to be extremely useful for analyzing binomial and Poisson data, response variables in the form of proportions or counts often exhibit more variation than expected from binomial or Poisson models. This chapter considers methods for estimation when overdispersion occurs, with an emphasis on the use of the double exponential family. The double exponential family is explored in chapter 5 as an estimation method for the measurement error problem. The topic of overdispersion is relevant to the discussion of measurement error for the following reason. As shown in section 3.4, if

$$E(y_i|x_i) = \mu_i = \beta_0 + \beta_1 x_i \text{ and}$$

$$\mathrm{Var}(y_i|x_i) = V(\mu_i)/m,$$

we know that

$$E(y_i|z_i) = \mu_i = \beta_0 + \beta_1 E(x_i|z_i) \text{ and}$$

$$\mathrm{Var}(y_i|z_i) = E[V(\mu_i)/m|z_i] + \beta_1^2 \mathrm{Var}(x_i|z_i).$$

From this, we might expect, at least as an approximation, that the distribution of y_i given z_i is similar to that of y_i given x_i, but with the first two moments as given above. If the distribution y_i given x_i is a one-parameter exponential family such as a binomial or

Poisson distribution, it is apparent that, given z_i, the distribution of y_i has greater variability than is explained by the binomial or Poisson model (although the Bernoulli model is an exception). The methods proposed for extra-binomial and extra-Poisson variation may be useful for the measurement error problem when the correct mean and variance for y_i given z_i are used as suggested in the techniques proposed by Prentice (1986) and Armstrong (1985) discussed in the previous chapter.

Section 4.1 of this chapter introduces overdispersion in its typical setting and illustrates problems that arise when overdispersion is ignored. Section 4.2 investigates likelihood-based solutions. Quasi-likelihood solutions are examined in section 4.3. Section 4.4 expands the concept of overdispersion to include covariates in the variance function. The last section in this chapter, section 4.5, introduces the use of Efron's (1986) double exponential family for dealing with overdispersion.

4.1 Introduction to Overdispersion

If an observed count is the sum of correlated Bernoulli trials or if counts for the same treatments have means which are random variables, there will be more variability than predicted by the binomial model. Likewise, in the Poisson setting, if the mean parameter is a random variable, the variability of the responses is greater than that predicted by the Poisson model. Methods for modeling overdispersion in binomial and Poisson models are introduced in the remaining portion of this section.

Although overdispersion has little effect on the estimation of mean parameters, Cox (1983) summarizes the problems that arise when overdispersion is ignored. The first is that the parameter estimates have a larger variance than anticipated by the simple model. The second problem is what Cox calls a possible loss of the efficiency of using statistics appropriate for the one-parameter family. In this case, using the customary fully efficient statistic for the one-parameter family may not accomplish full efficiency in the presence of overdispersion.

4.2 Fully Parametric Models for Overdispersion

4.2.1 The Correlated Bernoulli Model for Proportions

Greenwood (1949) dealt with the following situation. Let $y_i = (U_{i1}+\ldots+U_{im_i})/m_i$ where the U_{ij}'s are binary variables from a single group and share a common $p_i = \Pr(U_{ij}=1)$ and $\sigma_{ij}^2 = p_i(1-p_i)$ for all $j=1,\ldots,m_i$ and let pairs of binary observations within a group have common correlation, ρ_i. Given these assumptions, Greenwood showed that the response variable y_i has the following mean and variance

$$E(y_i) = p_i \quad \text{and}$$

$$\mathrm{Var}(y_i) = \sum_{j=1}^{m_i} \sigma_{ij}^2 + \sum_{j\neq k} \rho_i \sigma_{ij} \sigma_{ik} = \frac{p_i(1-p_i)}{m_i\theta_i},$$

where $\theta_i^{-1} = [1+(m_i-1)\rho_i]$. Greenwood asked Irwin (1954) to derive the distribution of y_i for this situation. Irwin accepted the challenge and showed that the correct distribution in this situation is the negative hypergeometric distribution which is more commonly known as the beta-binomial distribution. Maximum likelihood techniques can be used for the estimation of the parameters with the beta-binomial distribution used as the likelihood function.

4.2.2 The Beta-Binomial Model for Proportions

The beta-binomial distribution for proportion can also be derived in the following setting. Suppose $y_i = (U_{i1}+\ldots+U_{im_i})/m_i$ where U_{ij} are binary variables from a single group that share a common mean $P_i = \Pr[U_{ij}=1]$. Let P_i be a continuous random variable independently distributed over (0,1). A convenient distribution for P_i is a beta distribution, because it is flexible and has some desirable properties. This assumption on P_i leads to a beta-binomial distribution for the distribution of y_i which has the following variance:

$$\mathrm{Var}(y_i) = \frac{p_i(1-p_i)}{m_i\theta_i}, \text{ where } \theta_i^{-1} = [1+(m_i-1)\rho_i],$$

as in the previous section. Once again, maximum likelihood techniques can be used to estimate the parameters.

4.2.3 The Correlated Poisson Model for Counts

Irwin (1954) also derived a correlated Poisson model. The model is a Poisson-type model except for the correlation factor. Suppose that p_i is the probability that a single case occurs within any of the

m_i intervals of length h, but what occurs in one interval is not independent of what occurs in another interval. This results in the following mean-variance relationship,

$$E(y_i) = m_i p_i \text{ and}$$

$$Var(y_i) = m_i p_i + (m_i - 1)\rho m_i p_i = m_i p_i [1 + (m_i - 1)\rho]$$

Because it is thought that p_i is small and m_i is large, this could be modeled as a Poisson distribution with $\mu_i = m_i p_i$ except for the correction that occurs. Irwin discovered that the corresponding distribution for this correlated Poisson model is the negative binomial distribution. The negative binomial distribution for counts plays an analogous role to the beta-binomial distribution for proportions.

4.2.4 The Negative Binomial Model for Counts

Suppose that y_i given v_i is distributed Poisson(v_i). Suppose further that v_i is a random variable such that

$$E(v_i) = \mu_i \text{ and}$$

$$Var(v_i) = \sigma^2 \mu_i.$$

This leads to the following:

$$E(y_i) = E[E(y_i | v_i)] = \mu_i \text{ and}$$

$$\mathrm{Var}(y_i) = E[\mathrm{Var}(y_i|v_i)] + \mathrm{Var}[E(y_i|v_i)]$$

$$= \mu_i + \sigma^2\mu_i.$$

The term $\sigma^2\mu_i$ is extra-Poisson variation caused by the fact that v_i is a random variable. A convenient distribution for y_i results if v_i is assumed to have a gamma distribution which leads to negative binomial distribution for y_i. With these assumptions, maximum likelihood techniques can be used to estimate the parameters.

4.3 Quasi-likelihood Methods for Overdispersion

Quasi-likelihood methods (Wedderburn, 1974) only require assumptions about the mean and variance of the distribution rather than full distributional specifications as required for maximum likelihood estimation. When we are dealing with one-parameter exponential distributions, quasi-likelihood and maximum likelihood lead to the same estimates. The practical use of quasi-likelihood, however, has arisen from the ability to add an extra parameter to the binomial or Poisson variance as a simple model for overdispersion. Suppose that Y is a vector of responses which behaves similar to binomial or Poisson responses except that the variability is greater or less than expected. One way to model Y involves the following mean and variance:

$$E(\mathbf{Y}) = \boldsymbol{\mu} \text{ and}$$

$$\mathrm{Var}(\mathbf{Y}) = V(\boldsymbol{\mu})/\theta,$$

where values of θ between 0 and 1 are appropriate for modeling overdispersion and $V(\mu)$ is a positive semi-definite matrix whose elements are known functions of μ. The explanatory variable is involved in the model through $\mu = g^{-1}(x'\beta)$, where $x_i' = [1\ x_i]$. Let $D=d\mu/d\beta$ be full rank for all β, which ensures identifiability of the parameters. The quasi-likelihood function is defined as,

$$\partial Q(\mu;y)/\partial\mu = V^{-1}(\mu)(y-\mu).$$

The maximum quasi-likelihood equations used to solve for β are given by:

$$D'V^{-1}(\mu)(y-\mu) = 0.$$

The convenience of this method is that it is not necessary to know $Q(\mu;y)$ or θ to estimate β. The calculations can be carried out in GLIM (Baker and Nelder,1978).

Williams (1982) proposed two models for dealing with extra-binomial variation (overdispersion) in logistic regression. In the first model, he allowed the probabilities of successes for a group to be unobserved variables, P_i, independently distributed on (0,1) where only mean and variance assumptions are made about P_i, i.e.

$$E(P_i) = p_i \text{ and}$$

$$Var(P_i) = \rho\cdot p_i(1-p_i),$$

where $p_i^{-1} = 1 + \exp(-x'\beta)$. He assumed that conditional on $P_i=p_i^*$, y_i is binomial$(m_i,p_i^*)/m_i$. These assumptions on P_i and y_i lead to

$$E(y_i) = p_i \text{ and}$$

$$Var(y_i) = \theta_i^{-1} p_i(1-p_i)/m_i,$$

where $\theta_i^{-1} = 1+(m_i-1)\rho$. The knowledge of the mean-variance relationship of y_i can be used to determine the quasi-likelihood function to estimate the parameter β. This procedure assumes that ρ is known, which is not typical. Williams suggests estimating β using an estimate of ρ and then updating ρ before estimating β again. He also gives a goodness of fit test for ρ to indicate when the estimate of ρ is satisfactory.

A second model for overdispersion given by Williams (1982) has the logit of P_i varying about $x_i'\beta$ with constant variance σ^2, instead of varying about p_i with constant variance. Estimation for this model is accomplished by minimizing the weighted squared deviations between each y_i and its approximate expectation. If σ^2 is small, then

$$E(P_i) = p_i \cong [1 + \exp(-x'\beta)]^{-1} \text{ and}$$

$$Var(P_i) = \sigma^2 p_i^2 (1-p_i)^2.$$

Estimation can be accomplished for β as it was in the previous model, except that,

$$\theta_i^{-1} = 1 + \sigma^2(m_i-1)p_i^*(1-p_i^*)$$

where p_i^* is the value of p_i based on the last estimate of β and σ^2 needs to be estimated instead of ρ.

Breslow (1984) discussed extra-Poisson variation in log-linear models similarly to Williams. Instead of assuming v_i is distributed as a gamma distribution, Breslow only required mean and variance assumptions. Suppose y_i are observed counts which, given p_i (the unknown rates) and m_i, are distributed Poisson with $E(y_i|p_i) = m_i p_i$. Suppose also that the relationship between p_i and $x_i'\beta$ can be described with a log-linear model, i.e. $\log(p_i) = x_i'\beta + e_i$, where e_i has mean 0 and variance σ^2. This results in the following approximate GLM:

$$E(y_i) = \mu_i \cong \exp(\log m_i + x_i'\beta) \text{ and}$$

$$Var(y_i) \cong \mu_i + \sigma^2\mu_i.$$

If σ^2 is known, maximum quasi-likelihood estimates can be obtained in GLIM by using weights $\theta_i = (1 + \sigma^2\hat{\mu}_i)^{-1}$ and, if necessary, σ^2 can be estimated and updated through the iterative procedure. Breslow includes the GLIM macro to carry out this estimation.

4.4 Methods for Incorporating Covariates in the Variance Function

Two methods for using covariates in modeling the overdispersion are presented in this section. The first is suggested by Nelder and Pregibon (1987) using an extended quasi-likelihood function. The second, Prentice (1986), uses the beta-binomial distribution.

4.4.1 The Extended Quasi-likelihood Method

The extended quasi-likelihood method proposed by Nelder and Pregibon (1987) allows for the overdispersion parameter to be a function of covariates. The extended quasi-likelihood function for an individual is defined to be

$$Q^{+}(y_i,\mu_i) = -1/2\{\log[2\pi V(y_i)/\theta_i] + D(y_i,\mu_i)\theta_i\}$$

where $$D(y_i,\mu) = -2\{Q(y_i,\mu) - Q(y_i,y_i)\} = -2\cdot\int_{y_i}^{\mu} \frac{y_i-u}{V(u)}\,du$$

and θ_i is the dispersion parameter. The quasi-likelihood function, Q, is defined in section 4.3.

The following general model used in extended quasi-likelihood methods was given first by Pregibon (1984),

$$g(\mu_i) = x_i'\beta$$
$$h(\theta_i) = z_i'\alpha$$
$$\mathrm{Var}(y_i) = \theta_i^{-1}V(\mu_i).$$

In this model both the mean and overdispersion are functions of covariates. Estimation involves maximizing Q^{+}, where θ_i depends on unknown parameters. Iterative least square techniques are used by alternatively fixing μ and θ and updating the other parameter. This estimation can be carried out in GLIM using a macro.

4.4.2 The Extended Beta-Binomial Model

Prentice (1986) used the beta-binomial framework to model covariates in the dispersion parameter. He considered a special regression problem which occurs when measurement error in the explanatory variables is the sole source of extra-binomial variation.

In this case, regression models are induced for both $p(z_i)$ and $\theta(z_i)$. He embedded the resulting expressions for $E(y_i|z_i)$ and $Var(y_i|z_i)$ into a beta-binomial model so that maximum likelihood can be used for estimation. The likelihood function, $L(p,\theta)$, for a beta-binomial is used to estimate β. The score vector and information matrix are obtained through

$$\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial p}\frac{\partial p}{\partial \beta} + \frac{\partial L}{\partial \theta}\frac{\partial \theta}{\partial \beta}$$

and a similar expression for $-\frac{\partial^2 L}{\partial \beta \partial \beta'}$. These expressions, along with the first and second derivatives of $p(z_i)$ and $\theta(z_i)$ with respect to β, make it possible to carry out standard asymptotic likelihood inference on β.

4.5 The Double Exponential Family

In his motivation for the double exponential family, Efron (1986) described overdispersion as the clumping of observations which leads to a reduction in the number of independent observations, such as the number of Bernoulli trials going into an observed count. Efron used a reduction in the sample size from m_i to $m_i\theta_i$, where θ_i is between 0 and 1, and introduced the double exponential family as an extension of one-parameter exponential families to allow for this dispersion parameter. The double exponential family not only provides for the mean to vary independently of the variance, but allows for both the mean and the variance to be functions of the observed covariates.

Efron used the double exponential family as an estimation technique which accounts for overdispersion in a robust fashion. However, it seems that the framework is also suitable for modeling when a specific, covariate-dependent structure for the variance is assumed. One case with these characteristics is the regression of a response variable on an imprecisely measured covariate, as suggested at the beginning of this chapter. Included in section 4.5.1 is the definition of the double exponential family distribution. Section 4.5.2 covers the basic estimation procedure for this distribution and some properties of the double exponential family.

4.5.1 Definition of the Double Exponential Family

The double exponential family introduced by Efron (1986) allows for an extra parameter in one-parameter exponential families to account for the extra variability in the model that cannot be explained by the one-parameter distribution. When extra variability is present, the double exponential family distribution behaves similarly to the corresponding original one-parameter exponential distribution if the sample size is thought of as being reduced from m_i to $m_i\theta_i$.

A double exponential family density can be defined as follows. If $g_{\mu,m}(y)$ is an ordinary one-parameter exponential family of density functions, i.e.

$$g_{\mu,m}(y) = e^{m[\eta y-\psi(\mu)]}, \tag{4.1}$$

where $\mu = E(y)$, y is the natural statistic, η is the natural parameter, and $\psi(\mu)$ is the normalizing function so that the density integrates to 1, then the family of density functions

$$f_{\mu,\theta,m}(y) = c(\mu,\theta,m)\cdot\theta^{1/2}\{g_{\mu,m}(y)\}^{\theta}\cdot\{g_{y,m}(y)\}^{1-\theta}, \tag{4.2}$$

is called a double exponential family with parameters μ, θ, and m. The term $c(\mu,\theta,m)$ ensures that the function integrates to 1.

The binomial distribution is used as an example, as follows. If $g_{p,m}(y)$ is the probability function for a binomial proportion with probability p and index m, i.e.

$$g_{p,m}(y) = \begin{pmatrix} m \\ my \end{pmatrix}\cdot p^{my}\cdot(1-p)^{m-my},$$

where $y = 0,\ 1/m,\ 2/m, \ldots,\ 1,$ then

$$f_{p,\theta,m}(y) = c(p,\theta,m)\cdot\theta^{1/2}\{g_{p,m}(y)\}^{\theta}\cdot\{g_{y,m}(y)\}^{1-\theta},$$

is called a double binomial distribution.

4.5.2 Properties of the Double Exponential Family

Some properties of the double exponential family distribution are that,

1) $c(\mu_i,\theta_i,m_i) \doteq 1$ and

2) $E(y_i) \doteq \mu_i$ and $Var(y_i) \doteq \theta_i^{-1}V(\mu_i)$.

Efron shows that these approximations are nearly exact. This allows us to drop $c(\mu_i,\theta_i,m_i)$ from all subsequent uses of the double exponential family distribution in this work.

A feature of a double exponential family distribution is that the mean-variance relationship is the same as the one-parameter exponential model except for the additional factor, θ, in the variance. As Efron puts it, the double exponential family allows us to take the "quasi" out of "quasi-likelihood" because the mean-variance relationship is the same as what is used for quasi-likelihood but full likelihood techniques and properties may be used. Efron's idea was that θ_i could contain covariates and, in this sense, the idea is very closely related to the extended quasi-likelihood method (Nelder and Pregibon, 1987).

Some other facts presented by Efron (1986) include the following:

1) When θ_i and m_i are fixed, (4.2) is an exponential family distribution with index μ_i.

2) The density of (4.2) represents the same probability distribution as (4.1) except the sample size is adjusted from m_i to $m_i\theta_i$.

3) When μ_i and m_i are fixed, (4.2) is an exponential distribution with index θ_i and

$$D(y_i,\mu_i) \mathrel{\dot\sim} (1/\theta_i)\chi_1^2 \quad \text{as } m_i \rightarrow \infty,$$

where $D(y_i,\mu_i) = 2m_iE_{\mu_i,m_i}\{\log[g_{y_i,m_i}(x)/g_{\mu_i,m_i}(x)]\}$.

Estimation can be accomplished using the Newton-Raphson method to obtain maximum likelihood estimates. When the assumptions are met for the double exponential family, it is possible to get maximum likelihood estimates even in the presence of overdispersion.

If the double exponential distribution specifies the correct log-likelihood function, the estimator enjoys the usual properties of maximum likelihood estimators, i.e. it is consistent and asymptotically normal.

Chapter 5

Measurement Error Model Estimation Using the Double Exponential Family

The proposed method is introduced for the special case of the identity link and later expanded to include nonlinear links. In section 1 of this chapter the model is presented along with its assumptions. Section 2 discusses the role of Efron's (1986) double exponential family in the proposed method. The details of the estimation process are given in section 3, along with the properties of the estimator under ideal assumptions. Section 4 introduces three approximations to the proposed method that are considered in later sections. A special case, when the distribution of y_i given z_i is only an approximation, is considered in section 5. Section 6 covers specific examples using the identity link. Section 7 expands the concept to the general nonlinear link. Notes about the requirement that $E(x_i|z_i)$ and $Var(x_i|z_i)$ must be known are contained in section 8.

5.1 The Model and Assumptions

The following model is assumed for the relationship between the response y_i and the true explanatory variable, x_i:

$$E(y_i|x_i) = \mu_i = \beta_0 + \beta_1 x_i, \text{ and}$$

$$Var(y_i|x_i) = V(\mu_i)/m_i.$$

It is assumed that,

1) the distribution of y_i given x_i is a known member of a one

parameter exponential family;

2) $y_1, \ldots, y_n$ are independent, and

3) $f(y_i|x_i, z_i) = f(y_i|x_i)$.

An exact model for the measurement error is not required here, except that values of $E(x_i|z_i)$ and $Var(x_i|z_i)$ must be estimated in the first stage of the analysis. Two measurement error models are considered to illustrate this. The first is an additive model,

$$z_i = x_i + d_i,$$

with x_i and d_i independent. The second is a multiplicative model but reduces to an additive model when logarithms are applied, i.e.,

$$\log z_i = \log x_i + \log d_i.$$

In particular, for the additive model it is convenient to assume $x_i \sim N(\mu_x, \sigma_x^2)$ and $d_i \sim N(0, \sigma_d^2)$. With these assumptions the distribution of z_i is normal with mean μ_x and variance $\sigma_x^2 + \sigma_d^2$. The resulting conditional distribution of x_i given z_i has a mean and variance as follows:

$$E(x_i|z_i) = \mu_x + R(z_i - \mu_x) \text{ and}$$

$$Var(x_i|z_i) = \sigma_x^2(1-R),$$

where R is the reliability ratio, $\sigma_x^2/(\sigma_x^2+\sigma_d^2)$. Thus, the assumptions on x_i and d_i allow us to calculate the necessary values, $E(x_i|z_i)$ and $Var(x_i|z_i)$.

For the multiplicative model, it is convenient to assume $\log x_i \sim N(\mu_x, \sigma_x^2)$ and $\log d_i \sim N(0, \sigma_d^2)$. With these assumptions, the distribution of $\log z_i$ is normal with mean μ_x and variance $\sigma_x^2 + \sigma_d^2$. The resulting conditional distribution of $\log x_i$ given $\log z_i$ has a mean and variance as follows:

$$E(\log x_i | \log z_i) = \mu_x + R[\log(z_i) - \mu_x] \text{ and}$$

$$\mathrm{Var}(\log x_i | \log z_i) = \sigma_x^2(1-R).$$

where $R = \sigma_x^2/(\sigma_x^2+\sigma_d^2)$. The moment generating function for this conditional distribution can be used to show that:

$$E(x_i | z_i) = z_i^R \exp[(1-R)\mu_x + \tfrac{1}{2}(1-R)\sigma_x^2]$$

and,

$$E(x_i^2 | z_i) = z_i^{2R} \exp[2(1-R)\mu_x + 2(1-R)\sigma_x^2].$$

This leads to the following variance:

$$\begin{aligned}\mathrm{Var}(x_i | z_i) &= E(x_i^2 | z_i) - [E(x_i | z_i)]^2 \\ &= z_i^{2R} \exp[2(1-R)\mu_x + (1-R)\sigma_x^2]\{\exp[(1-R)\sigma_x^2] - 1\}\end{aligned}$$

These are two models for which $E(x_i | z_i)$ and $\mathrm{Var}(x_i | z_i)$ may be estimated with the available z_i's and some extra information about the distribution of the measurement error. Other methods, such as the use of a validation data set, may be used to estimate these. In what follows $E(x_i | z_i)$ and $\mathrm{Var}(x_i | z_i)$ are taken as known and the particular model for the distribution of x_i and the measurement error are not relevant.

5.2 The Double Exponential Family as an Approximation

The assumptions in the previous section lead to the following induced mean and variance for the conditional distribution of y_i given z_i:

$$E(y_i|z_i) = \mu_i^* = \beta_0 + \beta_1 \cdot E(x_i|z_i) \quad \text{and}$$

$$\begin{aligned} Var(y_i|z_i) &= \frac{1}{m_i} \cdot E[V(\mu_i)|z_i] + \cdot \beta_1^2 \cdot Var(x_i|z_i) \\ &= \frac{V(\mu_i^*)}{m_i \theta_i} \end{aligned}$$

where θ_i depends on the variance function $V(\cdot)$. For the primary cases of interest, when y_i given x_i is binomial or Poisson, $E[V(\mu_i)|z_i]$ can be easily expressed in terms of $E(x_i|z_i)$ and $Var(x_i|z_i)$. If the density function for the conditional distribution of y_i given z_i were available, estimates of the unknown parameters could be obtained by using maximum likelihood techniques. However, the density cannot usually be expressed in closed form. The idea is to use a density which approximates the distribution of y_i given z_i and has the same mean and variance as given above.

The proposed method uses Efron's (1986) double exponential family distribution to approximate the density for the induced model. This is closely related to Prentice's (1986) work in which the distribution of y_i given z_i is approximated by the beta-binomial distribution. These distributions are not believed to be the true distribution of y_i given z_i but are useful and convenient approximations for the purpose of obtaining estimates. By specifying the distribution, it is possible to use maximum likelihood estimation techniques to estimate the parameters of interest (although the resulting estimates only maximize a likelihood which approximates the true likelihood based on the distribution of y_i given z_i).

5.2.1 Introduction to the Proposed Method

The double exponential family distribution is based on a one-parameter exponential family distribution but contains an extra parameter, θ_i, which is used to model the dispersion. In the current problem, the parameter, θ_i, is a function of the measurement error variance and the unknown regression coefficients.

One feature of the double exponential family is that it allows us to model the dispersion parameter as a function of covariates. Using this feature the proposed method capitalizes on the fact that the variance of y_i given z_i is a function of β when the estimation of β is carried out with maximum likelihood techniques. The purpose of using information about β contained in the $\mathrm{Var}(y_i|z_i)$, is to more efficiently estimate β than is possible with, say, Armstrong's (1985) method. When information about the parameters contained in the variance is used in the estimation procedure, it is referred to as a <u>feedback model</u> (Carroll and Rupert,1982).

The proposed method is in the same spirit as the work proposed by Prentice (1986) and Armstrong (1985). Prentice's method is also a feedback method, but differs because the distribution of y_i given z_i is approximated by the beta-binomial distribution, instead of the double binomial distribution. Armstrong's method is not a feedback method, because it treats $Var(y_i|z_i)$ as known before each iteration and updates it after each iteration. His method assumes the same distributional form for the distribution of y_i given z_i as is thought to be true for the distribution of y_i given x_i, but accounts for the dispersion parameter in weights, w_i, and in the use of $E(x_i|z_i)$ as the explanatory variable rather than z_i. The use of the double binomial distribution was not intended to lead to "better" estimates than those from the beta-binomial distribution but the double exponential family approach is more flexible in that it could be used for the Poisson case as well.

5.2.2 Basis for Approximating the Distribution of y_i given z_i

As previously mentioned, the proposed method involves approximating the density of the induced model for y_i given z_i. The basis for this approximation includes the following reasons.

1) It is possible, in some situations, to write down an expression for the exact likelihood for the distribution of y_i given z_i with the given assumptions, but often it does not have a closed form solution. Thus, solving the exact maximum likelihood equations can be very difficult and involve numerical integration.

2) The double exponential family distribution offers a convenient framework for incorporating the mean-variance relationship given at the beginning of section 5.2, which is not substantially different from the original exponential family distribution assumed for y_i given x_i.

5.3 Details of Estimation and Properties of the Estimator

Recall that the induced mean and variance for this model are

$$E(y_i|z_i) = E[E(y_i|x_i)|z_i] = \beta_0 + \beta_1 \cdot E(x_i|z_i) = \mu_i^* \quad \text{and} \qquad (5.2)$$

$$\begin{aligned} Var(y_i|z_i) &= E[Var(y_i|x_i)|z_i] + Var[E(y_i|x_i)|z_i] \\ &= \frac{1}{m_i} E[V(\mu_i)|z_i] + \beta_1^2 \cdot Var(x_i|z_i). \qquad (5.3) \\ &= \frac{V(\mu_i^*)}{m_i \theta_i} \end{aligned}$$

It is assumed that both $E(x_i|z_i)$ and $Var(x_i|z_i)$ are known, or estimated in a first stage of the analysis as in the pseudo-likelihood approach used by Carroll et. al. (1984). The remaining parameters to be estimated in the induced model are $\beta = (\beta_0, \beta_1)$.

Let ℓ denote the log-likelihood based on the double exponential family distribution. The score statistic is obtained through

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \mu^*} \frac{\partial \mu^*}{\partial \beta} + \frac{\partial \ell}{\partial \theta} \frac{\partial \theta}{\partial \beta}$$

and a similar expression for the information matrix

$$I = \frac{\partial^2 \ell}{\partial \beta \partial \beta'}$$

$$= \frac{\partial}{\partial \beta}\left[\frac{\partial \ell}{\partial \mu^*}\frac{\partial \mu^*}{\partial \beta} + \frac{\partial \ell}{\partial \theta}\frac{\partial \theta}{\partial \beta}\right]$$

$$= \left[\frac{\partial}{\partial \beta}\left(\frac{\partial \ell}{\partial \mu^*}\right)\frac{\partial \mu^*}{\partial \beta'} + \frac{\partial \ell}{\partial \mu^*}\frac{\partial^2 \mu^*}{\partial \beta \partial \beta'}\right] + \left[\frac{\partial}{\partial \beta}\left(\frac{\partial \ell}{\partial \theta}\right)\frac{\partial \theta}{\partial \beta'} + \frac{\partial \ell}{\partial \theta}\frac{\partial^2 \theta}{\partial \beta \partial \beta'}\right]$$

$$= \left[\left(\frac{\partial^2 \ell}{(\partial \mu^*)^2}\frac{\partial \mu^*}{\partial \beta} + \frac{\partial^2 \ell}{\partial \theta \partial \mu^*}\frac{\partial \theta}{\partial \beta}\right)\frac{\partial \mu^*}{\partial \beta'} + \frac{\partial \ell}{\partial \mu^*}\frac{\partial^2 \mu^*}{\partial \beta \partial \beta'}\right] + \left[\left(\frac{\partial^2 \ell}{\partial \theta \partial \mu^*}\frac{\partial \mu^*}{\partial \beta} + \frac{\partial^2 \ell}{(\partial \theta)^2}\frac{\partial \theta}{\partial \beta}\right)\frac{\partial \theta}{\partial \beta'} + \frac{\partial \ell}{\partial \theta}\frac{\partial^2 \theta}{\partial \theta \partial \beta \partial \beta'}\right].$$

The estimation process is iterative, where β is updated at each step using the Newton-Raphson method as follows,

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - I^{-1} \cdot \frac{\partial \ell}{\partial \beta}$$

where $\hat{\beta}^{(t)}$ is the current estimate of β, $\hat{\beta}^{(t+1)}$ is the updated estimate, and I and $\frac{\partial \ell}{\partial \beta}$ are evaluated at $\hat{\beta}^{(t)}$.

If the double exponential family distribution specifies the correct log-likelihood function, then the estimator enjoys the properties of maximum likelihood estimators, i.e. it is consistent and asymptotically normal with mean β and variance I^{-1}. The simulation results seem to indicate the same properties hold true for the proposed estimator when the conditions of our example are satisfied. The details of the method are shown in section 5.6 on three special models for y_i given x_i: normal, binomial, and Poisson.

5.4 Approximations to the Proposed Method

The three approximations that are introduced in this section are modifications of the double exponential family, maximum likelihood method. The third approximation is considered as a special case in section 5.5. All of the approximations are illustrated on the chromosome aberration data example in section 6.1. The modifications are based on the extent to which information about β in the variance of y_i given z_i is used. Recall that

$$\mathrm{Var}(y_i|z_i) = \frac{V(\mu_i^*)}{m_i\theta_i},$$

where $\theta_i^{-1} = \frac{E[V(\mu_i)|z_i]}{V(\mu_i^*)} + m_i\beta_1^2\frac{\mathrm{Var}(x_i|z_i)}{V(\mu_i^*)}$ (this expression simplifies for the binomial and Poisson cases, in particular for the binomial case with identity link

$$\theta_i^{-1} = 1 + (m_i-1)\beta_1^2 v_i^*, \text{ where } v_i^* = \frac{\mathrm{Var}(x_i|z_i)}{V(\mu_i^*)}). \qquad (5.1).$$

The modifications involve replacing certain pieces of θ_i by their estimates and treating then as known. Armstrong's method, for example, involves the replacement of all of θ_i by its current estimates. For simplicity all the approximations are discussed in terms of the binomial identity link case, but it is very similar for the Poisson case and non-identity link.

The first approximation are referred to as approximate DEF method 1. In this method the only feedback information about β_1 in $\mathrm{Var}(y_i|z_i)$ from (5.1) that is used is contained in the β_1^2 term. The term V_i^* is estimated at each iteration of the Newton-Raphson method but treated as known. Two reasons for using this approximation include: 1) it is thought that most of the information about β_1 might be contained in the β_1^2 term and 2) the derivatives for the estimation of β are greatly simplified. If the estimates are not greatly affected by this approximation it may be preferred over using the complete feedback model.

The approximate DEF method 2 is an intermediate step between the first and third approximations. This method allows for β_1^2 to be replaced by another parameter, say α_1, in (5.1), but V_i^* is still treated as a function of β. This approximation allows for the feedback information about β in $\mathrm{Var}(y_i|z_i)$, except from the β_1^2 term.

Approximate DEF method 3 combines both of the first two approximations. This approximation allows for θ_i^{-1} to be independent of β, so that this method is no longer a feedback method, but more along the lines of Armstrong's (1985) method. In this approximation β_1^2 is replaced by α_1 and V_i^* is estimated at each iteration of the Newton-Raphson method but treated as known.

5.5 Some Asymptotic Results When the Double Exponential Family is Viewed as an Approximation

It is of interest to know how well the asymptotic properties of the maximum likelihood estimator apply when the correct distribution of y_i given z_i is not the double exponential family. Some of White's (1982) results concerning properties of maximum likelihood estimators based on incorrect likelihoods are relevant, but I have not been able to apply the theorems to the complete double exponential family estimator. The results below relate to the approximate DEF method 3 introduced in the section 5.4, with the restriction that the natural link is used. The results demonstrate the asymptotic properties of the double exponential family estimator when V_i^* is taken to be known. Additional results related to the small sample distribution of the double exponential family estimator are discussed in Chapter 7.

Suppose $g(y_i|z_i;\nu^o)$ is the true distribution of y_i given z_i with parameter vector $\nu^o=(\beta_0^o,\beta_1^o,\alpha^o)$. The distribution for y_i given z_i, on which the "maximum likelihood estimates" are obtained, is specified to be the general double exponential family distribution where θ_i^{-1} is independent of β_1 such that $\theta_i^{-1}= 1 + \alpha\cdot V_i^*$ with V_i^* known. This can be expressed in the general form with natural parameter, η_i, as

$$f(y_i|z_i;\nu) = \theta_i^{\frac{1}{2}}\left[\exp\left\{\frac{y_i\eta_i-b(\eta_i)}{w_i}+c(y_i)\right\}\right]^{\theta_i}\left[\exp\left\{\frac{y_i\tilde{\eta}_i-b(\tilde{\eta}_i)}{w_i}+c(y_i)\right\}\right]^{1-\theta_i}$$

$$= \theta_i^{\frac{1}{2}}\cdot\exp\left(\frac{\theta_i}{w_i}\left[[y_i\eta_i-b(\eta_i)]-[y_i\tilde{\eta}_i-b(\tilde{\eta}_i)]\right] + \frac{y_i\tilde{\eta}_i-b(\tilde{\eta}_i)}{w_i}+c(y_i)\right)$$

where $\nu=(\beta_0,\beta_1,\alpha)$ are the unknown parameters, $\eta_i=h(\mu_i^*)=\beta_0+\beta_1 z_i^*$ with $z_i^*=E(x_i|z_i)$, and $h(\cdot)$ is the natural link function where $\tilde{\eta}_i=h(y_i)$. The corresponding log-likelihood for y_i given z_i is

$$\ell(\nu;y_i,z_i)=\log[f(y_i|z_i;\nu)]=\tfrac{1}{2}\log(\theta_i)+\tfrac{1}{2}\theta_i D(y_i;\eta_i)+[y_i\tilde{\eta}_i-b(\tilde{\eta}_i)]/w_i+c(y_i)\Big\}$$

where $D(y_i;\eta_i)=\left[\frac{y_i\eta_i-b(\eta_i)}{w_i} - \frac{y_i\tilde{\eta}_i-b(\tilde{\eta}_i)}{w_i}\right]$ is the deviance. The existence of a unique maximum to the following expression is a necessary assumption for the use of White's theorem:

$$L = E_{\nu^o}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \ell(\nu;y_i,z_i)\right\} \tag{5.4}$$

where expectation is with respect to the true distribution, $g(y_i|z_i;\nu^o)$, and is denoted by the subscript ν^o. The maximization can be done by considering the following equations

$$\frac{\partial}{\partial\beta}L = 0 \quad \text{and} \quad \frac{\partial}{\partial\alpha}L = 0.$$

Subject to regularity conditions involving the exchangeability of derivatives and integrals, these equations reduce to

$$\frac{\partial}{\partial\beta}L=\sum_{i=1}^{n}\theta_i\frac{b(\eta_i^o)-b(\tilde{\eta}_i)}{w_i}\left(\frac{\partial\eta_i}{\partial\beta}\right) = 0 \quad \text{and}$$

$$\frac{\partial}{\partial\alpha}L=\sum_{i=1}^{n}\left[\frac{1}{2\theta_i} - \tfrac{1}{2}E_{\nu^o}[D(y_i;\eta_i)]\right]\frac{\partial\theta_i}{\partial\beta} = 0.$$

By using Taylor's first order approximations on $b(\tilde{\eta}_i)$ and $D(y_i;\eta_i)$ about η_i^o, these equations reduce to

$$\frac{\partial}{\partial\beta}L = \left[\sum_{i=1}^{n}\frac{\theta_i}{w_i}z_i z_i'\right](\hat{\beta}-\beta^o)=0 \quad \text{and}$$

$$\frac{\partial}{\partial\alpha}L=\sum_{i=1}^{n}\tfrac{1}{2}\left[\frac{V_i^*}{\hat{\theta}_i} - \frac{V_i^*}{\theta_i^o}\right] = 0,$$

where $\hat{\theta}_i^{-1} = 1 + \hat{\alpha}V_i^*$ and $Z_i' = [1\ E(x_i|z_i)]$. If $E(x_i|z_i)$ does not equal 0 or 1 for all i then the unique solution to these equations is $\hat{\nu} = \nu^{\circ}$. Thus, the solution to these equations converge to the true values as desired.

White defines the following matrices in order to arrive at the information matrix for the misspecified model:

$$A = E_{\nu^{\circ}}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \nu \partial \nu'}\right\} \text{ and } B = E_{\nu^{\circ}}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \nu}\left(\frac{\partial \ell}{\partial \nu}\right)'\right\}.$$

The variance-covariance matrix of the asymptotic distribution of $\hat{\nu}$ is defined as

$$C = A^{-1}BA^{-1}$$

evaluated at $\hat{\nu}=\nu^*$, where ν^* is the solution to (5.4). For our problem where the parameters of interest are the β's, the covariance matrix of interest becomes

$$C_{11} = A_{11}^{-1}B_{11}A_{11}^{-1}$$

where $A_{11} = E_{\beta^{\circ}}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \beta \partial \beta'}\right\}$ and $B_{11} = E_{\nu^{\circ}}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \beta}\left(\frac{\partial \ell}{\partial \beta}\right)'\right\}$.
When C_{11} is evaluated at $\beta=\beta^*$, the solution to (5.4), C_{11} reduces to A_{11} which is the usual asymptotic covariance matrix. The result can be expressed as

$$\sqrt{n}(\hat{\beta} - \beta^{\circ}) \xrightarrow{\mathcal{L}} N[0, C_{11}(\beta^{\circ})] \quad \text{as } n \longrightarrow \infty,$$

because β° is the solution for $\hat{\beta}$ in (5.4). Thus, in this special case the estimator is consistent and asymptotically normal with variance-covariance matrix as indicated by maximum likelihood theory as if the double exponential family were correct.

5.6 Special Cases with Identity Link

This section illustrates the proposed method using three special cases involving exponential family distributions: 1) normal with unknown variance; 2) binomial, and 3) Poisson. All of these cases use one of the measurement error models discussed in section 5.1, where it is assumed that $E(x_i|z_i)$ and $Var(x_i|z_i)$ are known or will be estimated in the first stage of the analysis and then treated as known in the estimation of β using the double exponential family.

5.6.1 The Normal Distribution

Suppose y_i can be modeled as follows:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

where $e_i \sim N(0,\sigma_e^2)$, and $\beta = (\beta_0, \beta_1)$ are the parameters of interest. Because $Var(y_i|x_i) = \sigma_e^2$, it follows from (5.2) and (5.3) that

$$E(y_i|z_i) = \beta_0 + \beta_1 E(x_i|z_i) = \mu_i^*$$

and

$$Var(y_i|z_i) = \sigma_e^2 + \beta_1^2 \cdot Var(x_i|z_i)$$

$$= 1/\theta_i$$

where $\theta_i^{-1} = \sigma_e^2 + \beta_1^2 \cdot Var(x_i|z_i)$. The normal model with unknown variance is a member of the double exponential family and may be written as

$$f_{\mu_i^*,\theta_i}(y_i|z_i)=\theta_i^{\frac{1}{2}}\left\{\frac{1}{\sqrt{2\pi}}\exp[-\tfrac{1}{2}(y_i-\mu_i^*)^2]\right\}^{\theta_i}\left\{\frac{1}{\sqrt{2\pi}}\exp[-\tfrac{1}{2}(y_i-y_i)^2]\right\}^{1-\theta_i}$$

$$= \frac{\theta_i^{\frac{1}{2}}}{\sqrt{2\pi}}\exp\left[-\frac{\theta_i}{2}(y_i-\mu_i^*)^2\right]. \tag{5.3}$$

The double exponential family approximation leads to the following for y_i given z_i:

$$y_i|z_i \sim N[\beta_0+\beta_1 E(x_i|z_i),\sigma_e^2+\beta_1^2 Var(x_i|z_i)]$$

This is exact if x_i given z_i is normal but an approximation otherwise.

If $Var(x_i|z_i)$ is constant, the maximum likelihood estimates are the least squares estimates in the simple linear regression of y_i on $E(x_i|z_i)$. If $Var(x_i|z_i)$ is not constant, there is information about β_1 in $Var(y_i|z_i)$. The corresponding log-likelihood for a single observation in terms of μ_i^* and θ_i is

$$\ell = -\tfrac{1}{2}\log(2\pi) + \tfrac{1}{2}\log(\theta_i) - \tfrac{1}{2}\theta_i(y_i-\mu_i^*)^2.$$

Estimation involves the score vector as described in section 5.3 and the following derivatives are needed:

$$\frac{\partial\ell}{\partial\beta_0}= \theta_i(y_i-\mu_i^*),$$

$$\frac{\partial \ell}{\partial \beta_1} = \theta_i(y_i-\mu_i^*)E(x_i|z_i) + \left[\tfrac{1}{2}(1/\theta_i) - \tfrac{1}{2}(y_i-\mu_i^*)^2\right] \frac{-2\beta_1 Var(x_i|z_i)}{[\sigma_e^2+\beta_1^2 Var(x_i|z_i)]^2},$$

The derivatives for the information matrix include the following,

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -\theta_i,$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -\theta_i E(x_i|z_i) + \left[(y_i-\mu_i^*)\right] \frac{-2\beta_1 Var(x_i|z_i)}{[\sigma_e^2+\beta_1^2 Var(x_i|z_i)]^2} \text{ and}$$

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = \{[-\theta_i E(x_i|z_i)+(y_i-\mu_i^*)\frac{\partial \theta}{\partial \beta_1}]E(x_i|z_i)\}$$
$$+ \{[(y_i-\mu_i^*)E(x_i|z_i)-\tfrac{1}{2}(1/\theta_i^2)\frac{\partial \theta}{\partial \beta_1}]\frac{\partial \theta}{\partial \beta_1} + \left[\tfrac{1}{2}(1/\theta_i) - \tfrac{1}{2}(y_i-\mu_i^*)^2\right]\frac{\partial^2 \theta}{\partial \beta_1^2}\},$$

where $\frac{\partial \theta}{\partial \beta_1} = \frac{-2\beta_1 Var(x_i|z_i)}{[\sigma_e^2+\beta_1^2 Var(x_i|z_i)]^2}$ and $\frac{\partial^2 \theta}{\partial \beta_1^2} = -2Var(x_i|z_i)\frac{[\sigma_e^2-3\beta_1^2 Var(x_i|z_i)]}{[\sigma_e^2+\beta_1^2 Var(x_i|z_i)]^3}$.

The above equations used with the Newton-Raphson method below yields the maximum likelihood solutions for β,

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - I^{-1}\cdot\frac{\partial \ell}{\partial \beta}.$$

Some simplification is possible if the method of scoring is used for the normal situation; but not for the binomial and Poisson cases that follow.

The purpose of the double exponential family is to be able to apply the same desirable properties of the normal distribution to distributions like the binomial and Poisson distribution. The next two subsections give an example of these two distributions.

5.6.2 The Binomial Case with Identity Link Function

Suppose $y_i|x_i$ is distributed binomial $(m_i,p_i)/m_i$, where

$$E(y_i|x_i) = p_i = \beta_0+\beta_1 x_i \text{ and}$$

$$\mathrm{Var}(y_i|x_i) = \frac{p_i(1-p_i)}{m_i}.$$

These assumptions lead to the following mean and variance for y_i given z_i:

$$E(y_i|z_i) = \beta_0+\beta_1 E(x_i|z_i) = p_i^* \text{ and}$$

$$\mathrm{Var}(y_i|z_i) = \frac{1}{m_i}E\{(\beta_0 + \beta_1 x_i)[1-(\beta_0 + \beta_1 x_i)]|z_i\} + \beta_1^2\cdot \mathrm{Var}(x_i|z_i)$$

$$= \frac{1}{m_i}p_i^*(1-p_i^*) + \left(1-\frac{1}{m_i}\right)\beta_1^2\cdot \mathrm{Var}(x_i|z_i)$$

$$= \frac{p_i^*(1-p_i^*)}{m_i\theta_i},$$

where $\theta_i^{-1} = \phi_i = 1+\dfrac{(m_i-1)\beta_1^2 \mathrm{Var}(x_i|z_i)}{p_i^*(1-p_i^*)}$. The mean and variance are in the same form as the original distribution of y_i given x_i with the addition of θ_i in the variance. This is the situation that the double exponential family is designed to handle. The double binomial distribution is defined to be:

$$f_{p_i^*,\theta_i}(y_i|z_i)=\theta_i^{\frac{1}{2}}\left[p_i^{*\,m_i y_i}(1-p_i^*)^{m_i-m_i y_i}\right]^{\theta_i}\left[y_i^{m_i y_i}(1-y_i)^{m_i-m_i y_i}\right]^{1-\theta_i}\binom{m_i}{m_i y_i}$$

and can be used to approximate the distribution of y_i given z_i. Using $f_{p_i^*,\theta_i}$ as the likelihood, the usual asymptotic inference can be performed. The corresponding log-likelihood for a single observation in terms of p_i^* and θ_i is

$$\ell = \tfrac{1}{2}\log(\theta_i) + m_i\theta_i[y_i\log(p_i^*)+(1-y_i)\log(1-p_i^*)] + m_i(1-\theta_i)[y_i\log(y_i)+(1-y_i)\log(1-y_i)] + \log\binom{m_i}{m_iy_i}.$$

Estimation involves the score vector as described in section 5.3 and the following derivatives are needed:

$$\frac{\partial\ell}{\partial\beta_0} = \frac{m_i\theta_i(y_i-p_i^*)}{p_i^*(1-p_i^*)}(1) + \frac{\partial\ell}{\partial\theta}\frac{\partial\theta}{\partial\phi}\frac{\partial\phi}{\partial\beta_0} \quad \text{and}$$

$$\frac{\partial\ell}{\partial\beta_1} = \frac{m_i\theta_i(y_i-p_i^*)}{p_i^*(1-p_i^*)}E(x_i|z_i) + \frac{\partial\ell}{\partial\theta}\frac{\partial\theta}{\partial\phi}\frac{\partial\phi}{\partial\beta},$$

where $\frac{\partial\ell}{\partial\theta} = \frac{1}{2}(1/\theta_i) + m_i[y_i\log(p_i^*)+(1-y_i)\log(1-p_i^*)] - m_i[y_i\log(y_i)+(1-y_i)\log(1-y_i)]$,

$$\frac{\partial\theta}{\partial\phi} = \frac{-1}{(\phi_i)^2}, \quad \frac{\partial\phi}{\partial\beta_0} = \frac{-(m_i-1)\beta_1^2\mathrm{Var}(x_i|z_i)[1-2\beta_0-2\beta_1E(x_i|z_i)]}{[p_i^*(1-p_i^*)]^2} \quad \text{and}$$

$$\frac{\partial\phi}{\partial\beta_0} = \frac{(m_i-1)\mathrm{Var}(x_i|z_i)[2\beta_0\beta_1-2\beta_0^2\beta_1-2\beta_0\beta_1^2E(x_i|z_i)+\beta_1^2E(x_i|z_i)]}{[p_i^*(1-p_i^*)]^2}.$$

The derivatives for the information matrix include the following,

$$\frac{\partial^2\ell}{\partial\beta_0^2} = \left[\left(\frac{\partial^2\ell}{(\partial p^*)^2}+\frac{\partial^2\ell}{\partial\theta\partial p^*}\frac{\partial\theta}{\partial\phi}\frac{\partial\phi}{\partial\beta_0}\right)\right]+\left[\left(\frac{\partial^2\ell}{\partial\theta\partial p^*}+\frac{\partial^2\ell}{(\partial\theta)^2}\frac{\partial\theta}{\partial\phi}\frac{\partial\phi}{\partial\beta_0}\right)\frac{\partial\theta}{\partial\phi}\frac{\partial\phi}{\partial\beta_0}+\frac{\partial\ell}{\partial\theta}\frac{\partial^2\theta}{(\partial\phi)^2}\frac{\partial^2\phi}{(\partial\beta_0)^2}\right],$$

$$\frac{\partial^2 \ell}{\partial\beta_0\partial\beta_1} = \left[\left\langle \frac{\partial^2 \ell}{(\partial p^*)^2} + \frac{\partial^2 \ell}{\partial\theta\partial p^*}\frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_0}\right\rangle E(x_i|z_i)\right] + \left[\left\langle \frac{\partial^2 \ell}{\partial\theta\partial p^*} + \frac{\partial^2 \ell}{(\partial\theta)^2}\frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_0}\right\rangle \frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_1} + \frac{\partial\ell}{\partial\theta}\frac{\partial^2\theta}{(\partial\psi)^2}\frac{\partial^2\psi}{\partial\beta_0\partial\beta_1}\right]$$

and

$$\frac{\partial^2 \ell}{\partial\beta_1^2} = \left[\left\langle \frac{\partial^2 \ell}{(\partial p^*)^2}E(x_i|z_i) + \frac{\partial^2 \ell}{\partial\theta\partial p^*}\frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_1}\right\rangle E(x_i|z_i)\right] + \left[\left\langle \frac{\partial^2 \ell}{\partial\theta\partial p^*}E(x_i|z_i) + \frac{\partial^2 \ell}{(\partial\theta)^2}\frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_1}\right\rangle \frac{\partial\theta}{\partial\psi}\frac{\partial\psi}{\partial\beta_1} + \frac{\partial\ell}{\partial\theta}\frac{\partial^2\theta}{(\partial\psi)^2}\frac{\partial^2\psi}{(\partial\beta_1)^2}\right],$$

where $\dfrac{\partial^2 \ell}{(\partial p^*)^2} = \dfrac{-m_i\theta_i[y_i - 2y_ip_i^* + (p_i^*)^2]}{[p_i^*(1-p_i^*)]^2}$, $\qquad \dfrac{\partial^2 \ell}{\partial\theta\partial p^*} = \dfrac{m_i(y_i - p_i^*)}{p_i^*(1-p_i^*)}$,

$$\frac{\partial^2 \ell}{(\partial\theta)^2} = \frac{-1}{(\theta_i)^2}, \qquad \frac{\partial^2\theta}{(\partial\psi)^2} = \frac{2}{\psi_i^3},$$

$$\frac{\partial^2\psi}{(\partial\beta_0)^2} = \frac{2(m_i-1)\beta_1^2 \mathrm{Var}(x_i|z_i)\{p_i^*(1-p_i^*) + [1-2\beta_0-2\beta_1E(x_i|z_i)]^2\}}{[p_i^*(1-p_i^*)]^3}$$

$$\frac{\partial^2\psi}{\partial\beta_0\partial\beta_1} = -2(m_i-1)\mathrm{Var}(x_i|z_i)\left(\frac{\{[\beta_1 - 2\beta_0\beta_1 - 3\beta_1^2E(x_i|z_i)]p_i^*(1-p_i^*)}{[p_i^*(1-p_i^*)]^3} - \frac{E(x_i|z_i)\beta_1^2[1-2\beta_0-2\beta_1E(x_i|z_i)]^2}{[p_i^*(1-p_i^*)]^3}\right)$$

and

$$\frac{\partial^2\psi}{(\partial\beta_1)^2} = 2(m_i-1)\mathrm{Var}(x_i|z_i)\left(\frac{\{[\beta_0 - \beta_0^2 - 2\beta_0\beta_1E(x_i|z_i) + \beta_1E(x_i|z_i)]p_i^*(1-p_i^*)}{[p_i^*(1-p_i^*)]^3} - \frac{E(x_i|z_i)[1-2\beta_0-2\beta_1E(x_i|z_i)][2\beta_0\beta_1 - 2\beta_0^2\beta_1 - 2\beta_0\beta_1^2E(x_i|z_i) + \beta_1^2E(x_i|z_i)]}{[p_i^*(1-p_i^*)]^3}\right).$$

The exact distribution of y_i given z_i, assuming the additive measurement error model, can be expressed as follows:

$$f(y_i|z_i) = \int_0^{\infty} f(y_i|x_i)\cdot f(x_i|z_i)dx_i$$

$$= \int_0^{\infty} \binom{m_i}{m_i y_i} p_i^{*\,m_i y_i}(1-p_i^*)^{m_i - m_i y_i}\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\tfrac{1}{2}(x_i-\mu)^2/\sigma^2]dx_i$$

where $\mu = \mu_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_d^2}(z_i-\mu_x)$ and $\sigma^2 = \frac{\sigma_x^2\cdot\sigma_d^2}{\sigma_x^2+\sigma_d^2}$. The exact distribution of y_i given z_i, as illustrated here, does not have a closed form solution. It can be approximated using numerical integration. The derivatives for this process can be very difficult to calculate and numerical integration can be complicated. Also, the actual calculation process is very time consuming even with the aid of a computer. Although the derivatives are difficult for the double binomial distribution, numerical integration is avoided and the computations are much easier. As it turns out, in the simulation for the example considered, the exact maximum likelihood is slightly less biased than the proposed method, with a slightly smaller mean square error. Very little, if any, information is lost by using the double binomial distribution instead of the exact method in the situations considered.

5.6.3 The Poisson Case with Identity Link Function

Suppose $y_i|x_i$ is distributed Poisson(μ_i), where

$$E(y_i|x_i) = \mu_i = \beta_0+\beta_1 x_i \text{ and}$$

$$\mathrm{Var}(y_i|x_i) = \mu_i.$$

Then, from (5.2) and (5.3)

$$E(y_i|z_i) = \beta_0+\beta_1 E(x_i|z_i) = \mu_i^* \text{ and}$$

$$\mathrm{Var}(y_i|z_i) = E[(\beta_0 + \beta_1 x_i)|z_i] + \beta_1^2 \cdot \mathrm{Var}(x_i|z_i)$$

$$= \mu_i^* + \beta_1^2 \cdot \mathrm{Var}(x_i|z_i)$$

$$= \mu_i^*/\theta_i$$

where $\theta_i^{-1} = 1 + \dfrac{\beta_1^2 \cdot \mathrm{Var}(x_i|z_i)}{\mu_i^*}$. The mean and variance are in the same form as the original distribution of y_i given x_i with the addition of θ_i in the variance. The double Poisson distribution used as an approximation to the distribution y_i given z_i is:

$$f_{p_i,\theta_i}(y_i|z_i) = \theta_i^{\frac{1}{2}} \cdot \left[\frac{e^{-\mu_i^*} \cdot \mu_i^{*y_i}}{y_i!}\right]^{\theta_i} \left[\frac{e^{-y_i} \cdot y_i^{y_i}}{y_i!}\right]^{1-\theta_i}.$$

Using $f_{\mu_i^*,\theta_i}$ as the likelihood, the usual asymptotic inference can be performed. The exact distribution of y_i given z_i, as it was for the binomial case in section 5.4.2, can be expressed as follows:

$$f(y_i|z_i) = \int_0^\infty \frac{e^{-\mu_i^*} \cdot \mu_i^{*y_i}}{y_i!} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\tfrac{1}{2}(x_i-\mu)^2/\sigma^2]\,dx_i$$

where $\mu = \mu_x + \dfrac{\sigma_x^2}{\sigma_x^2+\sigma_d^2}(z_i-\mu_x)$ and $\sigma^2 = \dfrac{\sigma_x^2 \cdot \sigma_d^2}{\sigma_x^2+\sigma_d^2}$.

The exact distribution of y_i given z_i as illustrated here does not have a closed form expression. It can be approximated using numerical integration. So the same type of advantages, for using the double Poisson distribution over the exact distribution, hold for the Poisson case as they did for the binomial case of the last section.

5.7 Special Cases with Nonlinear Link

This section features the proposed method using nonlinear link functions. In the case of the identity link, the assumption that $E(x_i|z_i)$ and $Var(x_i|z_i)$ are known resulted in β being the only unknown part of $E(y_i|z_i)$ and $Var(y_i|z_i)$. Thus, for the nonlinear link,

$$g(\mu_i) = \beta_0+\beta_i x_i,$$

$E(y_i|z_i) = E[g^{-1}(\beta_0+\beta_1 x_i|z_i]$ so either additional assumptions are needed or $g^{-1}(\cdot)$ must be approximated so that $E(x_i|z_i)$ and $Var(x_i|z_i)$ provide the necessary information. If an attempt is made to linearize $g^{-1}(\cdot)$ in x_i, a first order Taylor's approximation about $E(x_i|z_i)$ yields

$$g^{-1}(\beta_0+\beta_1 x_i) \doteq g^{-1}[\beta_0+\beta_1 E(x_i|z_i)] + \dot{g}^{-1}[\beta_0+\beta_1 E(x_i|z_i)]\beta_1[x_i-E(x_i|z_i)],$$

where $\dot{g}^{-1}(\cdot)$ is the derivative of $g^{-1}(\cdot)$ with respect to its argument. When this approximation is used the mean and variance for the distribution of y_i given z_i are as follows:

$$E(y_i|z_i) = E[E(y_i|x_i)|z_i] = E[g^{-1}(\beta_0+\beta_1 x_i)|z_i] \doteq g^{-1}[\beta_0+\beta_1 E(x_i|z_i)]$$

and

$$
\begin{aligned}
\mathrm{Var}(y_i|z_i) &= E[\mathrm{Var}(y_i|x_i)|z_i] + \mathrm{Var}[E(y_i|x_i)|z_i] \\
&= E[\mathrm{Var}(y_i|x_i)|z_i] + \mathrm{Var}[g^{-1}(\beta_0+\beta_1 x_i)|z_i] \\
&\doteq E[\mathrm{Var}(y_i|x_i)|z_i] + \mathrm{Var}[g^{-1}[\beta_0+\beta_1 E(x_i|z_i)] \\
&\qquad + \dot{g}^{-1}[\beta_0+\beta_1 E(x_i|z_i)]\beta_1[x_i-E(x_i|z_i)]|z_i] \\
&\doteq E[\mathrm{Var}(y_i|x_i)|z_i] + \{\dot{g}^{-1}[\beta_0+\beta_1 E(x_i|z_i)]\}^2\beta_1^2\mathrm{Var}(x_i|z_i)
\end{aligned}
$$

In both of the examples considered later in this section, $\mathrm{Var}(y_i|z_i)$ is a function of $E(x_i|z_i)$ and $\mathrm{Var}(x_i|z_i)$ when the first order Taylor's approximation is used. Now assuming $E(x_i|z_i)$ and $\mathrm{Var}(x_i|z_i)$ are known, and because we know $g^{-1}(\cdot)$ and $\dot{g}^{-1}(\cdot)$, all that remains to estimate is the parameters of interest, as is true for the identity link.

Two cases are considered, the binomial distribution and the Poisson distribution. In section 5.5.1, the binomial case using the natural link function, the logit link, is investigated. Section 5.5.2 uses the log link for the Poisson case. Once again, either of the measurement error models described in section 5.1 would be appropriate assumptions for the nonlinear link situation.

5.7.1 The Binomial Case with Logit Link Function

Suppose $y_i|x_i$ is distributed binomial $(m_i, p_i)/m_i$ where

$$E(y_i|x_i) = p_i = g^{-1}(\beta_0+\beta_1 x_i) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \quad \text{and}$$

$$\mathrm{Var}(y_i|x_i) = \frac{p_i(1-p_i)}{m_i},$$

where $g(p_i) = \mathrm{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0+\beta_1 x_i$ and $\beta = (\beta_0, \beta_1)$ are the parameters of interest. These assumptions lead to the following mean and variance for y_i given z_i:

$$E(y_i|z_i) = E[g^{-1}(\beta_0+\beta_1 x_i)|z_i], \text{ and}$$

$$\mathrm{Var}(y_i|z_i) = E\left[\left.\frac{p_i(1-p_i)}{m_i}\right| z_i\right] + \mathrm{Var}[g^{-1}(\beta_0 + \beta_1 x_i)|z_i]$$

$$= \frac{1}{m} E\{g^{-1}(\beta_0+\beta_1 x_i)[1-g^{-1}(\beta_0+\beta_1 x_i)]|z_i\} + \mathrm{Var}[g^{-1}(\beta_0+\beta_1 x_i)|z_i]$$

When the Taylor's approximation is used, the mean and variance for the distribution of y_i given z_i are as follows:

$$E(y_i|z_i) \doteq g^{-1}[\beta_0+\beta_1 E(x_i|z_i)] = p_i^* \quad \text{and}$$

$$\mathrm{Var}(y_i|z_i)=\frac{1}{m_i}E\{g^{-1}(\beta_0+\beta_1x_i)[1-g^{-1}(\beta_0+\beta_1x_i)]|z_i\}+\mathrm{Var}[g^{-1}(\beta_0+\beta_1x_i)|z_i]$$

$$\doteq \frac{1}{m_i}\Big[g^{-1}[\beta_0+\beta_1E(x_i|z_i)]-\{g^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2 - \beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)\Big] + \beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)$$

$$\doteq \frac{1}{m_i}\Big[g^{-1}[\beta_0+\beta_1E(x_i|z_i)]-\{g^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\Big] + \left(1-\frac{1}{m_i}\right)\beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)$$

$$\doteq \frac{1}{m_i}p_i^*(1-p_i^*) + \left(1-\frac{1}{m_i}\right)\beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)$$

$$\doteq \frac{p_i^*(1-p_i^*)}{m_i\theta_i},$$

where $$\theta_i^{-1}= 1 + \frac{(m_i-1)\beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)}{p_i^*(1-p_i^*)}$$

$$= 1 + (m_i-1)\beta_1^2p_i^*(1-p_i^*)\mathrm{Var}(x_i|z_i).$$

The simplification in θ_i results from the fact that for the logit link

$$p_i^*(1-p_i^*) = \dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)].$$

Now $E(y_i|z_i)$ and $\mathrm{Var}(y_i|z_i)$ are functions only of the unknown parameters $\beta = (\beta_0,\beta_1)$. The mean and variance are in the same form as the original distribution of y_i given x_i with the addition of θ_i in the variance. A double binomial model with this mean and variance can be used to estimate β. The exact distribution of y_i given z_i does not have a closed form solution and must be approximated using numerical integration.

5.7.2 The Poisson Case with Log Link

Suppose $y_i|x_i$ is distributed Poisson(μ_i), where

$$E(y_i|x_i) = \mu_i = g^{-1}(\beta_0+\beta_1 x_i) = \exp(\beta_0+\beta_1 x_i) \text{ and}$$

$$Var(y_i|x_i) = \mu_i,$$

where $g(\mu_i) = \log(\mu_i) = \beta_0+\beta_1 x_i$ is the natural log link and $\beta = (\beta_0,\beta_1)$ are the parameters of interest. These assumptions lead to the following mean and variance for y_i given z_i:

$$E(y_i|z_i) = E[g^{-1}(\beta_0+\beta_1 x_i)|z_i] \quad \text{and}$$

$$Var(y_i|z_i) = E(\mu_i|z_i) + Var[g^{-1}(\beta_0 + \beta_1 x_i)|z_i]$$

$$= E[g^{-1}(\beta_0+\beta_1 x_i)|z_i] + Var[g^{-1}(\beta_0+\beta_1 x_i)|z_i].$$

When the Taylor series expansion is used, the mean and variance for the distribution of y_i given z_i are as follows:

$$E(y_i|z_i) \doteq g^{-1}[\beta_0+\beta_1 E(x_i|z_i)] = \mu_i^* \quad \text{and}$$

$$\begin{aligned}\mathrm{Var}(y_i|z_i) &= E[g^{-1}(\beta_0+\beta_1x_i)|z_i] + \mathrm{Var}[g^{-1}(\beta_0+\beta_1x_i)|z_i]\\ &\doteq g^{-1}[\beta_0+\beta_1E(x_i|z_i)] + \beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)\\ &\doteq \mu_i^* + \beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)\\ &\doteq \frac{\mu_i^*}{\theta_i},\end{aligned}$$

where $\theta_i^{-1} = 1 + \beta_1^2\{\dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]\}^2\mathrm{Var}(x_i|z_i)/\mu_i^*$

$$= 1 + \beta_1^2\cdot\mu_i^*\cdot\mathrm{Var}(x_i|z_i).$$

The simplification in θ_i results from the fact that for the log link $\mu_i^* = \dot{g}^{-1}[\beta_0+\beta_1E(x_i|z_i)]$. Now $E(y_i|z_i)$ and $\mathrm{Var}(y_i|z_i)$ are functions only of the unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)$. The mean and variance are in the same form as the original distribution of y_i given x_i with the addition of θ_i in the variance. A double Poisson model with this mean and variance can be used to estimate $\boldsymbol{\beta}$. The only change from the Poisson case with the identity link is the form of μ_i^*. The exact distribution of y_i given z_i does not have a closed form solution and must be approximated using numerical integration.

5.8 Notes about the Requirement that $E(x|z)$ and $\mathrm{Var}(x|z)$ are Known

Recall from section 2.2 that even with the assumption that the vector (y_i, x_i, d_i) was distributed normally, all of the parameters were not identifiable. In this particular example, we had one more parameter than minimal sufficient statistics. The need for additional information in that section took a number of different forms, one of which was the variance of the measurement error.

The means for both x_i and z_i can be estimated from the observed z_i since the expected value of the z_i is μ_x. The variance of the observed z_i is equal to $\sigma_x^2 + \sigma_d^2$. If either σ_x^2 or σ_d^2 are known we could estimate the other value by subtracting the known value from σ_z^2. Knowledge of σ_d^2 provides the necessary information for estimating $E(x_i|z_i)$ and $Var(x_i|z_i)$, with the assumptions we are using for the measurement error models introduced in section 5.1.

It is not likely that such knowledge of the measurement error variance is available. A sensitivity analysis involves fixing the measurement error variance at various possible values in order to estimate β, and then exploring the results of interest over a plausible range of these values.

Another alternative to arrive at the information needed for estimating $E(x_i|z_i)$ and $Var(x_i|z_i)$ involves approximating $E(x_i|z_i)$ by a linear function of z_i. This idea is explored by Cochran (1975) for the additive measurement error model and by Schafer (1988) for the multiplicative measurement error model. Thus, by assuming $E(x_i|z_i)$ can be expressed as a linear function of z_i, say $L_0 + L_1 z_i$, it is possible to solve for L_0 and L_1 in terms of known or estimable values.

Chapter 6

Data Examples

This chapter focuses on the analysis of the examples introduced in chapter 1. Although complete analysis of these data sets are not given, estimates of parameters are provided using the proposed double exponential family estimator and several alternative methods. Section 6.1 contains the model and resulting analysis for the chromosome aberration data. Section 6.2 considers an example of the association between serum cholesterol and heart disease.

6.1 Chromosome Aberration Data Analysis

Recall from section 1.2.1 that this data set involves 649 survivors of the atomic bomb dropped on Hiroshima. The underlying model used for this data set results from research conducted at the Radiation Effects Research Foundation in Hiroshima and those affiliated with this organization, including Gilbert (1984). The underlying model is laid out in the following points.

1) The response variable, y_i, is the proportion of cells with chromosome aberrations, out of 100 examined, and the explanatory variable, x_i, is an individual's true radiation exposure. The distribution of y_i given x_i is

$$\text{binomial}(100,p_i)/100.$$

2) The relationship between p_i and x_i is linear, i.e.

$$p_i = \beta_0 + \beta_1 x_i.$$

This assumption is rather unusual for the binomial framework, but there are biological reasons as well as empirical evidence for this model. (See Otake and Prentice (1984) for a more detailed discussion.)

3) The observed measurements of radiation, z_i, are thought to be related to the true radiation by

$$z_i = x_i \cdot d_i$$

The values of $E(x_i|z_i)$ and $Var(x_i|z_i)$ are taken from work done by Pierce, Stram, and Vaeth (1989) based on the assumptions that $x_i \sim$ Weibull and $d_i \sim$ lognormal. The standard deviation of d_i is assumed here to be 0.30; as is assumed by Prentice (1986). It would be wise to explore the sensitivity of parameter estimates to this choice of measurement error standard deviation and to the particular model from which $E(x_i|z_i)$ and $Var(x_i|z_i)$ are derived, but this will not be pursued here. Because the measurement error is believed to be multiplicative, $Var(x_i|z_i)$ is a function of z_i. The values of $E(x_i|z_i)$ and $Var(x_i|z_i)$ that are used for the chromosome aberration data set are in Table 1.

4) The data has been grouped into the following seven dose categories measured in rads: (0, 1-99, 100-199, 200-299, 300-399, 400-499, 500 and above). The values that are used for z_i in each group

Table 1 -- Conditional Means and Variances used for x_i given z_i.

Measured Radiation, z	E(x|z)	Var(x|z)
0.0	0.0	0.0
38.0	38.35	133.9
143.9	135.60	1596.4
244.1	220.40	4133.8
346.9	303.11	7763.3
440.7	375.59	17789.6
666.6	541.22	23872.8

Measurement Error = 0.30.

are given in the first column of Table 1. These values are the average of the observations within the group.

Recall from section 5.5.2 that the mean and variance for y_i given z_i, when the distribution of y_i given x_i is binomial with an identity link function, are

$$E(y_i|z_i) = \beta_0 + \beta_1 E(x_i|z_i) = p_i^* \quad \text{and}$$

$$\mathrm{Var}(y_i|z_i) = \frac{p_i^*(1-p_i^*)}{m_i\theta_i},$$

where $\theta_i^{-1} = 1 + (m_i-1)\beta_1^2 V_i^*$ with $V_i^* = \dfrac{\mathrm{Var}(x_i|z_i)}{p_i^*(1-p_i^*)}$. This mean-variance relationship is incorporated in some way into all the methods illustrated in the data example that follows, except for the naive method which ignores measurement error.

The methods that were used to analyze the data include the following,

1) Naive method --- This method uses ordinary binomial regression treating the z_i's as if they were the true radiation exposures.

2) <u>Naive method corrected for attenuation</u> --- Measurement error spreads out the explanatory variable, attenuating the slope parameter. By dividing the naive estimate by

$$L_1 = E(d_i) \cdot \frac{var(x_i)}{var(z_i)}$$

a consistent estimate of β is obtained. This result is for the multiplicative measurement error model given by Schafer (1988).

3) <u>Naive method with z replaced by E(x|z)</u> --- This method uses ordinary binomial regression (maximum likelihood) treating $E(x_i|z_i)$ as if it is the true radiation exposures but makes no adjustment to the variance function.

4) <u>IWLS</u> --- Armstrong's method is exactly the same as using iteratively weighted least square (IWLS) in the regression of y_i on $E(x_i|z_i)$ with weights, $1/Var(y_i|z_i)$, updated at each iteration. (See section 3.3.4 for a more thorough discussion of Armstrong's method.)

5) <u>Prentice's method</u> --- This method uses the beta-binomial distribution to approximate the density of y_i given z_i. Maximum likelihood techniques are used to estimate β. (See section 3.3.5 for a more thorough discussion of Prentice's method.)

6) <u>Double Exponential Family (DEF) method</u> --- This method uses Efron's double binomial distribution to approximate the density of y_i given z_i. Maximum likelihood techniques are then used to estimate β. (See section 5.3 for more details on estimation using the double exponential family distribution.)

7) <u>Approximate DEF method 1</u>--- In this method, V_i^* is estimated at each iteration but treated as known. Maximum likelihood estimates are obtained from the DEF model with

$$\theta_i^{-1} = 1 + (m_i-1)\beta_1^2 V_i^*,$$

where V_i^* is taken as known.

8) <u>Approximate DEF method 2</u>--- This method replaces β_1^2 with a different parameter, but V_i^* is still a function of β. The form of θ_i^{-1} becomes

$$\theta_i^{-1} = 1 + (m_i-1)\alpha_1 V_i^*,$$

where V_i^* is not assumed known. Estimation is carried out as in 6) above.

9) <u>Approximate DEF method 3</u>--- This method allows for the variance to be independent of β. The form of θ_i^{-1} becomes

$$\theta_i^{-1} = 1 + (m_i-1)\alpha_1 V_i^*,$$

where V_i^* is treated as known in each iteration as in 7) above.

10) <u>Lognormal MLE</u> -- This method involves numerical integration of the exact distribution of y_i given z_i assuming that $\log(x_i)$ given $\log(z_i)$ is normal with an estimated mean and variance for the distribution of $\log(x_i)$. Numerical integration is used to approximate the derivative values and the Newton-Raphson method is used to estimate the parameters. It is not expected that the data follow these distributional assumptions but there is some

interest in exploring the results on this oversimplified model, as used by Prentice (1986).

Table 2 contains estimates of the variance of y_i given z_i separated, for each radiation group of the chromosome data, as follows

$$\mathrm{Var}(y_i|z_i) = \frac{p_i^*(1-p_i^*)}{m_i} + \left(1 - \frac{1}{m_i}\right)\beta_1^2 \mathrm{Var}(x_i|z_i)$$

$$= \begin{pmatrix}\text{usual}\\ \text{binomial}\\ \text{variation}\end{pmatrix} + \begin{pmatrix}\text{extra-}\\ \text{binomial}\\ \text{variation}\end{pmatrix}.$$

Notice that there is very little overdispersion due to measurement

Table 2 -- Contributions to $\mathrm{Var}(x_i|z_i)$.

Methods	Measured Radiation, z, in hundreds						
Variance Contributions	0	0-1	1-2	2-3	3-4	4-5	5+
Naive							
Binomial variation	8	18	43	65	87	105	145
Regress on E(x|z)							
Binomial variation	8	18	45	67	87	103	138
IWLS							
Binomial variation	8	18	45	67	87	104	138
Extra-variation	0	1	13	35	65	98	200
Prentice							
Binomial variation	7	19	46	69	90	107	143
Extra-variation	0	1	15	39	72	109	221
DEF							
Binomial variation	7	19	49	73	96	114	151
Extra-variation	0	1	17	45	83	127	257
DEF Approx. 1							
Binomial variation	7	20	52	78	102	121	159
Extra-variation	0	1	20	52	97	147	299
DEF Approx. 2							
Binomial variation	8	19	45	67	87	104	138
Extra-variation	0	1	13	35	65	98	200
DEF Approx. 3							
Binomial variation	8	19	45	67	86	103	137
Extra-variation	0	1	13	34	64	97	196

Table values have been multiplied by 10^5.

error in the low dose categories but that it accounts for over half of the variation in the two high dose groups.

Table 3 contains the resulting estimates for the slope parameter. The following items are technical comments about the estimates:

1) The naive estimate of slope is the smallest, as expected, since measurement error, in general, tends to attenuate the regression.

2) The Lognormal MLE method assumes, along with multiplicative measurement error, that the d_i's and x_i's are distributed lognormal which implies that the z_i's are lognormal. An examination of the distribution of the z_i's reveals that this is not true. Also these assumptions do not allow for zeros, although there are many present in this example. Making this incorrect assumption for this data, we observe similar results to those when it is assumed that the z_i's are measured without error.

Table 3 -- Slopes for Chromosome Aberration Data Analysis with Measurement Error = 0.30

Methods	Estimate of Slope ($x10^6$)	SE of Estimate ($x10^6$)
naive	252	5.7
Corrected for Att.	288	6.5
Regress on $E(x\|z)$	290	6.5
IWLS	291	8.4
Prentice	306	8.5
DEF	330	9.2
DEF (Approx. 1)	356	10.4
DEF (Approx. 2)	291	14.7
DEF (Approx. 3)	288	14.8
Exact MLE	256	7.5

3) The IWLS, Prentice and DEF methods produced very similar estimates of β_1, with the DEF method having a slightly higher estimated standard error.

4) The IWLS method, which uses no information about β_1 contained in the variance to estimate β_1, seems to produce similar results to the approximate DEF method 2 that replaces β_1^2 with α_1. This seems to imply that as far as feedback information in the variance is concerned, the β_1^2 term contains most of this information (and not V_i^*).

6.2 Heart Disease Data Analysis

This data was introduced in section 1.2.2. A goal of this study was to measure the degree of association for serum cholesterol level and heart disease for 350,977 men between the ages of 35 and 57. It should be pointed out that it would be quite important to have information on age and smoking habits to analyze this data. The data reported publicly by Iso et al. (1989) contains age-adjusted rates of heart disease. These will be used to illustrate the double exponential family method. Figure 2, on page 6, suggests that the log of $E(y_i)$ is approximately linear in observed cholesterol. The problem once again for this data set is that we only have the observed cholesterol level which contains measurement error. In addition to the presence of this form of measurement error, the data is also grouped into the following 10 categories, (less then 140, 140-159, 160-179, 180-199, 200-219, 220-239, 240-259, 260-279, 280-299, more than 300).

Let the response for individual k, y_k, be defined as follows,

$$y_k = \begin{cases} 1 & \text{if the kth individual had coronary heart disease} \\ 0 & \text{otherwise} \end{cases}$$

and let x_k = true serum cholesterol for the kth individual, and suppose

$$E(y_k|x_k) = p_k = \exp(\beta_0+\beta_1 x_k).$$

Let z_k be the observed level of cholesterol and suppose that the requirements for the normal additive measurement error model of section 5.1 are satisfied so that

$$x_k|z_k \sim N(\mu_x(1-R)+Rz_k,\ R\sigma_d^2). \tag{6.1}$$

Recall that $R = \sigma_x^2/(\sigma_x^2+\sigma_d^2)$. This assumption implies that z_k is normally distributed. Figure 3 shows a frequency polygon for the distribution of observed cholesterol levels and a superimposed grouped normal distribution. The assumption seems to be reasonable, although the distribution is slightly skewed to the right.

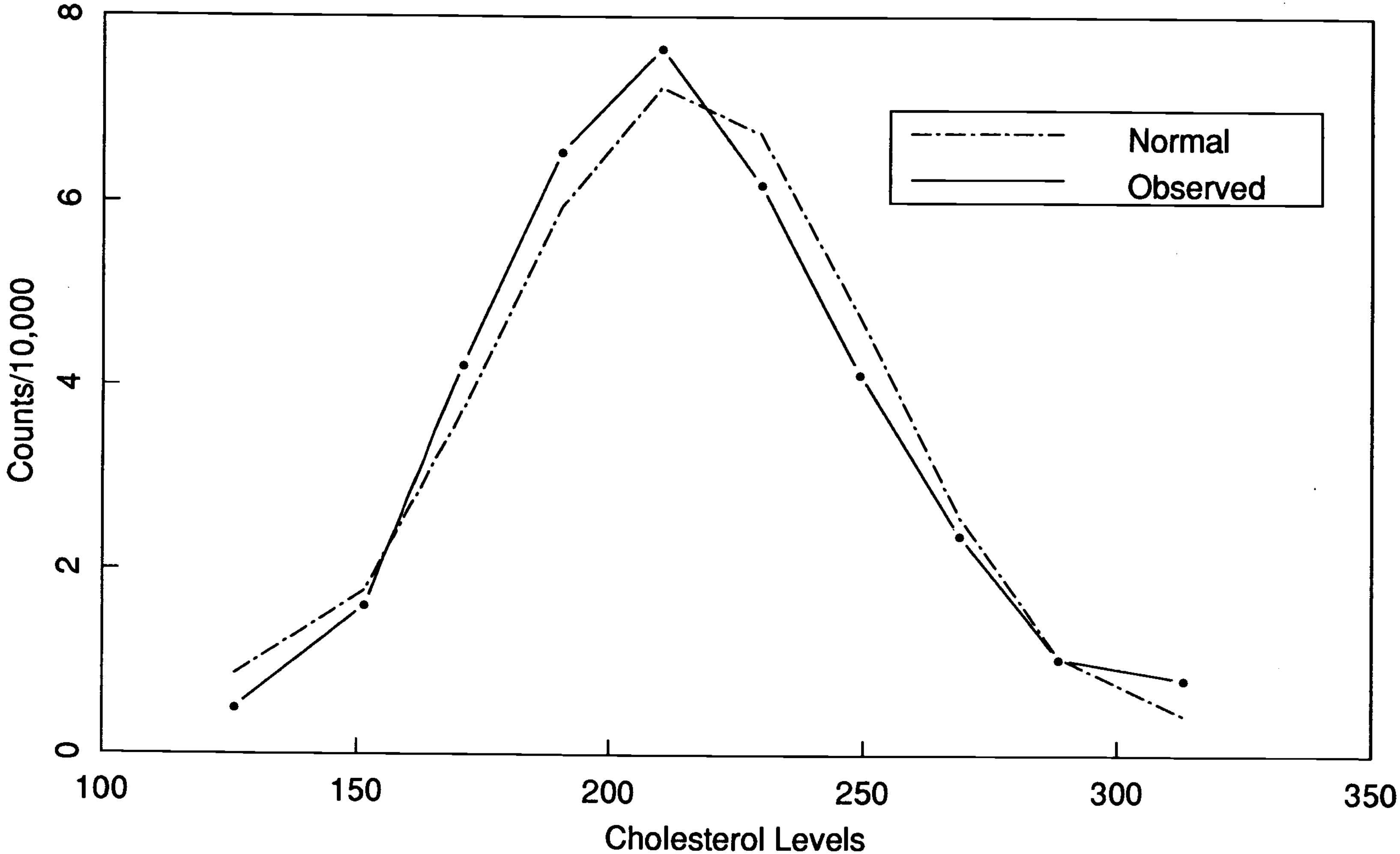
The mean for the distribution of y_k given z_k is

$$\begin{aligned} E(y_k|z_k) &= E[\exp(\beta_0+\beta_1 x_k)|z_k] \\ &\doteq \exp[\beta_0'+\beta_1 E(x_k|z_k)] = p_k' \end{aligned}$$

where $\beta_0' = \beta_0 + \frac{1}{2}\beta_1^2\sigma_d^2 R$. The moment generating function of (6.1) may be used to evaluate $E[\exp(\beta_1 x_k)|z_k]$. The corresponding variance is

$$\mathrm{Var}(y_k|z_k) = p_k'(1-p_k')$$

Figure 3. Frequency Polygon for Cholesterol Levels



because y_k is a Bernoulli random variable.

The following discussion concerns the mean and variance of observed counts in intervals grouped by measured cholesterol level. Let y_i^* denote the proportion with coronary heart disease in group i, i.e.

$$y_i^* = \left(\sum_{\Delta_i} y_k\right) / m_i^*, \text{ where } \Delta_i = \{k : c_{i-1} < z_k < c_i\}$$

and m_i^* equals the number at risk in group i, for i=1,...,10. Let $z_i^* = E(z_k | c_{i-1} < z_k < c_i)$. Because z_k is assumed normal z_i^* is the mean of a truncated normal. Table 4 contains the values z_i^*, using formulas for the mean and variance of a truncated normal distribution taken from Johnson and Kotz (1984).

Table 4 -- Means and Variances for Truncated Normal

Serum Cholesterol mg/dl	Group (i)	$E(z_k \| c_{i-1} < z_i < c_i)$	$Var(z_k \| c_{i-1} < z_i < c_i)$
< 140	1	125.6	165.68
140-159	2	151.4	30.40
160-179	3	171.0	31.58
180-199	4	190.6	32.33
200-219	5	210.1	32.66
220-239	6	229.7	32.54
240-259	7	249.2	31.98
260-279	8	268.8	30.99
280-299	9	288.3	29.60
> 300	10	313.2	142.42

The distribution of y_i^* given z_i^* has the following mean and variance when a first order Taylor's approximation is used for $\exp(\beta_0^*+\beta_1 Rz_k)$ and $\exp[2(\beta_0^*+\beta_1 Rz_k)]$,

$$E(y_i^*|z_i^*) = \sum_{\Delta_i} E(y_k/m_i^*|z_i^*)$$

$$= \exp(\beta_0^* + \beta_1 Rz_i^*) = p_i^* \quad \text{and}$$

$$\mathrm{Var}(y_i^*|z_i^*) = \sum_{\Delta_i} \mathrm{Var}(y_k/m_i^*|z_i^*)$$

$$= \frac{p_i^*(1-p_i^*)}{m_i^*} + \frac{p_i^{*2}}{m_i^*}\beta_1^2R^2\mathrm{Var}(z_k|c_{i-1}<z_k<c_i)$$

$$\doteq \frac{p_i^*}{m_i^*} + \frac{p_i^{*2}}{m_i^*}\beta_1^2R^2\mathrm{Var}(z_k|c_{i-1}<z_k<c_i)$$

$$\doteq \frac{p_i^*}{m_i^*\theta_i},$$

where $\theta_i^{-1} = 1 + \beta_1^2R^2p_i^*\mathrm{Var}(z_k|c_{i-1}<z_k<c_i)$, $\beta_0^* = \beta_0+\frac{1}{2}\sigma_d^2\beta_1^2R+\beta_1\mu_x(1-R)$, and $p_i^* \doteq p_i^*(1-p_i^*)$, because of the assumptions $(1-p_i^*) \doteq 1$ when p_i^* is small.

Based on what has already been discussed in this section the following model will be fitted,

$$E(y_i|z_i^*) = p_i^* = \exp(\alpha_0+\alpha_1 z_i^*) \quad \text{and}$$

$$\mathrm{Var}(y_i|z_i^*) = \frac{p_i^*}{m_i^*\theta_i},$$

where $\alpha_0=\beta_0^*$, $\alpha_1= \beta_1 R$ and $\theta_i^{-1}=1+\alpha_1^2 p_i^* \mathrm{Var}(z_k|c_{i-1} < z_k < c_i)$.

In order to estimate β, estimates of the nuisance parameters, μ_x, σ_x^2 and σ_d^2 are needed. The mean of x_k, μ_x, is estimated using the sample mean of z_i^*. The variance of z_k is estimated with the variances given in Table 4 using formulas from the truncated normal via Johnson and Kotz (1984). All that remains is to estimate σ_d^2. A profile likelihood analysis is used to estimate σ_d^2 and a value of $\sigma_d^2 = 38.1$ is estimated. Ideally, it is desired to have more knowledge about the measurement error associated with serum cholesterol.

The estimation of the parameters is done using a program written in Pascal. Appendix B contains the code for a similar program used for the binomial logit link model. Table 5 contains the estimates of the slopes using three methods: the naive method using z_i^*; iteratively weighted least squares, and the double exponential family, maximum likelihood, method.

Table 5 -- Slopes for Heart Disease Data with $\sigma_d^2 = 38.1$

Methods	Estimate of Slope ($\times 10^4$)	SE of Estimate ($\times 10^4$)
naive using z*	119	5.5
IWLS	119	5.5
DEF	122	5.6

Technical comments for the heart disease data:

1) Because the values of p_i^* are small the values of $\theta_i \doteq 1$ (i.e. very little overdispersion due to measurement error).
2) The DEF method is affected more by changes in σ_d^2 than the other methods presented here.

This is not a complete analysis of the heart disease data. The above discussion does, however, illustrates how the double exponential family method can be used to estimate β in similar situations.

Chapter 7

Comparison of the Double Exponential Family Method With Other Methods

The first section of this chapter discusses some conceptual issues distinguishing the proposed method, based on the double exponential family, from Prentice's (1986) approach and iteratively weighted least squares (IWLS). The latter is basically Armstrong's method and has been suggested by others including Pierce, Stram and Vaeth (1989). The remainder of the chapter compares these estimators and the exact maximum likelihood estimator, using simulations for a model in which y_i given x_i is binomial and the measurement error is multiplicative. Section 7.2 addresses the comparisons of efficiency of these methods when the underlying model is binomial with identity link (similar to the chromosome aberration data) and with logit link. The issue of robustness is considered for two situations in section 7.3. Section 7.4 presents the simulation results for the binomial model with identity link, when sample estimates of the nuisance parameters are used. In section 7.5, the standard errors for the methods are compared to the sampling standard deviations.

The five methods compared in the simulation are: 1) the naive method; 2) the MLE based on a multiplicative measurement error model with lognormal x_i and lognormal measurement errors (referred to as the Lognormal MLE); 3) iteratively weighted least squares (IWLS); 4) Prentice's method, and 5) the double exponential family (DEF) method. A summary of these methods is given in section 6.1.

7.1 Conceptual Differences

The proposed method is similar in spirit to that suggested by Prentice (1986), who approximated the density of y_i given z_i by a beta-binomial distribution. In fact, the estimation procedure is identical except for the approximating distribution. The methods are both feedback methods, meaning information about β in the variance of y_i given z_i is used in the estimation of the β.

The proposed method offers two advantages over Prentice's method. First it is more general, working just as easily with any one-parameter exponential family as it does with a binomial distribution. Second, it is a simpler approach which directly incorporates the correct mean-variance relationship. An additional weakness of using the beta-binomial distribution is that the convergence time for this method is more than ten times longer than any of the other methods considered, except the Lognormal MLE method. On the other hand, for proportions, the beta-binomial offers a richer class of models to approximate the correct density of y_i given z_i.

The biggest difference between the proposed method and IWLS is that the latter is not a feedback estimator. The IWLS method updates the weight at each stage of the iteration using current estimates and does not use information in the observed variability concerning the regression parameters of interest. The advantage of IWLS is that it is a simple method and it should be preferred if it can be shown to be relatively efficient. A further issue concerning the use of the feedback methods is robustness. Because information about β in

$\mathrm{Var}(y_i|z_i)$ is used in these procedures, they will be more susceptible to outliers.

7.2 Efficiency

For each of the situations of interest, 300 data sets are generated. For each of these data sets an estimate of slope, β_1, is calculated using each of the five methods listed at the beginning of this chapter. The particular data sets act as blocks and the treatments of interest are the methods of estimation. The question of efficiency is addressed by comparing the exact maximum likelihood estimator (MLE) to the other estimators in terms of bias and mean square error (MSE). Because it is known that the exact MLE is asymptotically efficient this estimator serves as a standard in order to judge how well the other estimators are performing.

In order to obtain estimates of bias and MSE, the following mixed model for the relative bias of the slope for each method over the different data sets is used:

$$\frac{(\hat{\beta}_{ik}-\beta_1)}{\beta_1} = \alpha_k + B_i + \epsilon_{ik}, \quad \text{for } i=1,\ldots,300 \text{ and } k=1,\ldots 5,$$

where

$\hat{\beta}_{ik}$ is the estimate of slope for the kth method on the ith data set,

β_1 is the true slope used in the simulation,

α_k is the bias of the kth method, as a proportion of β_1,

B_i is the blocking effect of the ith data set with $B_i \sim N(0,\sigma_b^2)$ and

ϵ_{ik} is the random error with $\epsilon_{ik} \sim N(0,\sigma_k^2)$.

Thus, the relative bias, α_k, is a fixed effect and the data sets are chosen at random so that the block effect, B_i, is a random effect. Relative MSE is estimated from this as follows,

$$E\left[\frac{(\hat{\beta}_{ik}-\beta_1)}{\beta_1}\right]^2 = \frac{\text{MSE}}{\beta_1^2} \doteq \hat{\alpha}_k^2 + \hat{\sigma}_b^2 + \hat{\sigma}_k^2,$$

7.2.1 Binomial Model with the Identity Link Function

7.2.1.1 Design of the Simulation

This simulation study is designed to recreate the chromosome aberration data as closely as possible. Three values of measurement error standard deviation, σ_d, (0.1, 0.3, 0.5), are considered. Three hundred data sets were generated for each value of σ_d. Each of these data sets contains the following elements for i=1,...,100:

1) $\log x_i \sim \text{normal}(\mu_x, \sigma_x^2)$ with μ_x=4.0 and σ_x^2=0.5,
2) $\log d_i \sim \text{normal}(0, \sigma_d^2)$,
3) $z_i = x_i \cdot d_i$,
4) $y_i \sim \text{bin}(100, p_i)/100$, where $p_i = 0.018 + 0.0003x_i$ and
5) $E(x_i|z_i)$ and $\text{Var}(x_i|z_i)$ are computed exactly, according to the lognormal distribution of x_i given z_i (see page 48 for details):

$$E(x_i|z_i) = z_i^R \exp[(1-R)\mu_x + \tfrac{1}{2}(1-R)\sigma_x^2] \quad \text{and}$$

$$\text{Var}(x_i|z_i) = z_i^{2R} \exp[2(1-R)\mu_x + (1-R)\sigma_x^2]\{\exp[(1-R)\sigma_x^2]-1\},$$

where $R = \sigma_x^2/(\sigma_x^2+\sigma_d^2)$.

In the chromosome aberration data, the standard deviation of the measurement error, σ_d, is thought to be about .3 as reported by Gilbert (1984). A measurement error value of .5 is thought to be extreme, but is included to judge the effect of an extremely large multiplicative measurement error. All the methods investigated reduce to the usual binomial regression estimates in the absence of measurement error. This is very evident for $\sigma_d = 0.1$, as the estimates are very close for all methods considered.

7.2.1.2 Results of the Simulation

The estimators of β_1 (the slope parameter) are compared according to relative bias and relative mean square error (MSE/β_1^2). The nuisance parameters μ_x, σ_x^2, and σ_d^2 are taken to be known in this study. The effect of estimating these is exploded in section 7.4. Appendix B contains the computer program for the simulation presented in section 7.2.2, which is very similar in structure and content to the program for this simulation. Only values where methods converged are included. As it turns out this is only a problem for the situation where the underlying model for log d_i is contaminated normal.

Figures 4, 5 and 6 contain pair-wise scatterplots of the 300 values of the four estimators of interest (Lognormal MLE, IWLS, Prentice and DEF) for measurement error standard deviations of .1, .3, and .5 respectively. The histograms for the bias as a proportion of β_1 are shown for each estimator individually. All four of the methods

Figure 4. Relative Bias of the Four Methods (Identity Link and Measurement Error Standard Deviation = 0.1)

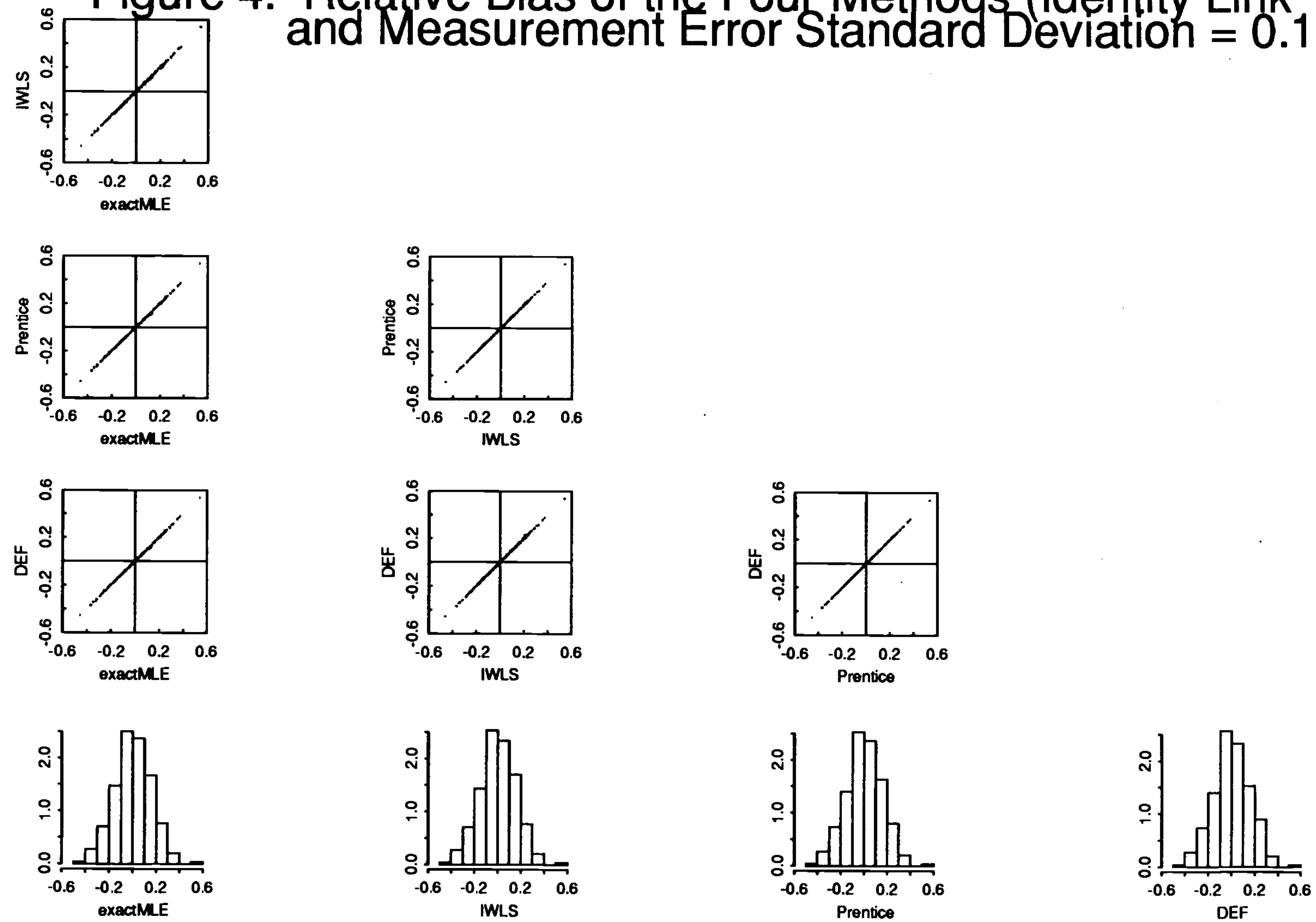


Figure 5. Relative Bias of the Four Methods (Identity Link and Measurement Error Standard Deviation = 0.3)

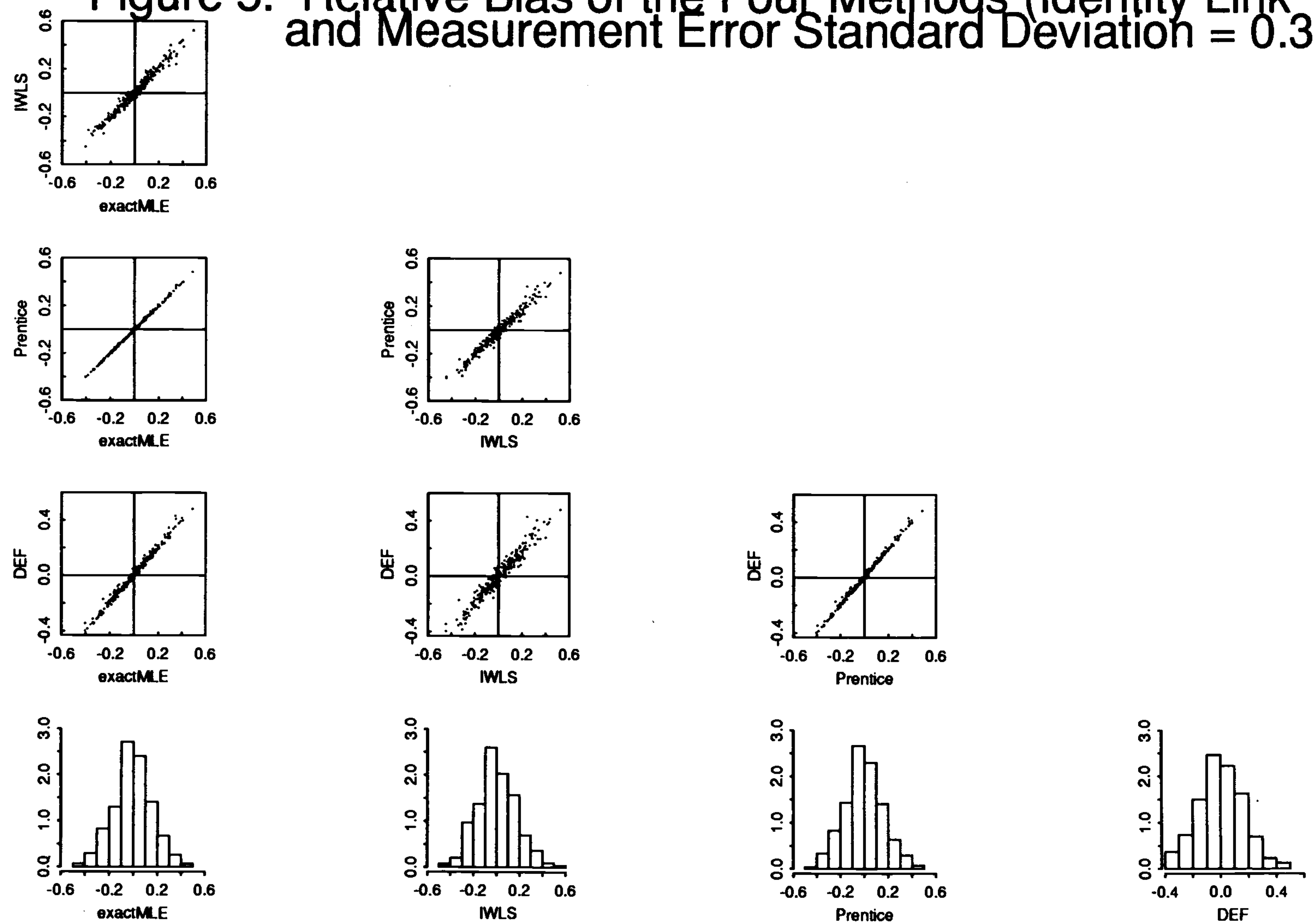
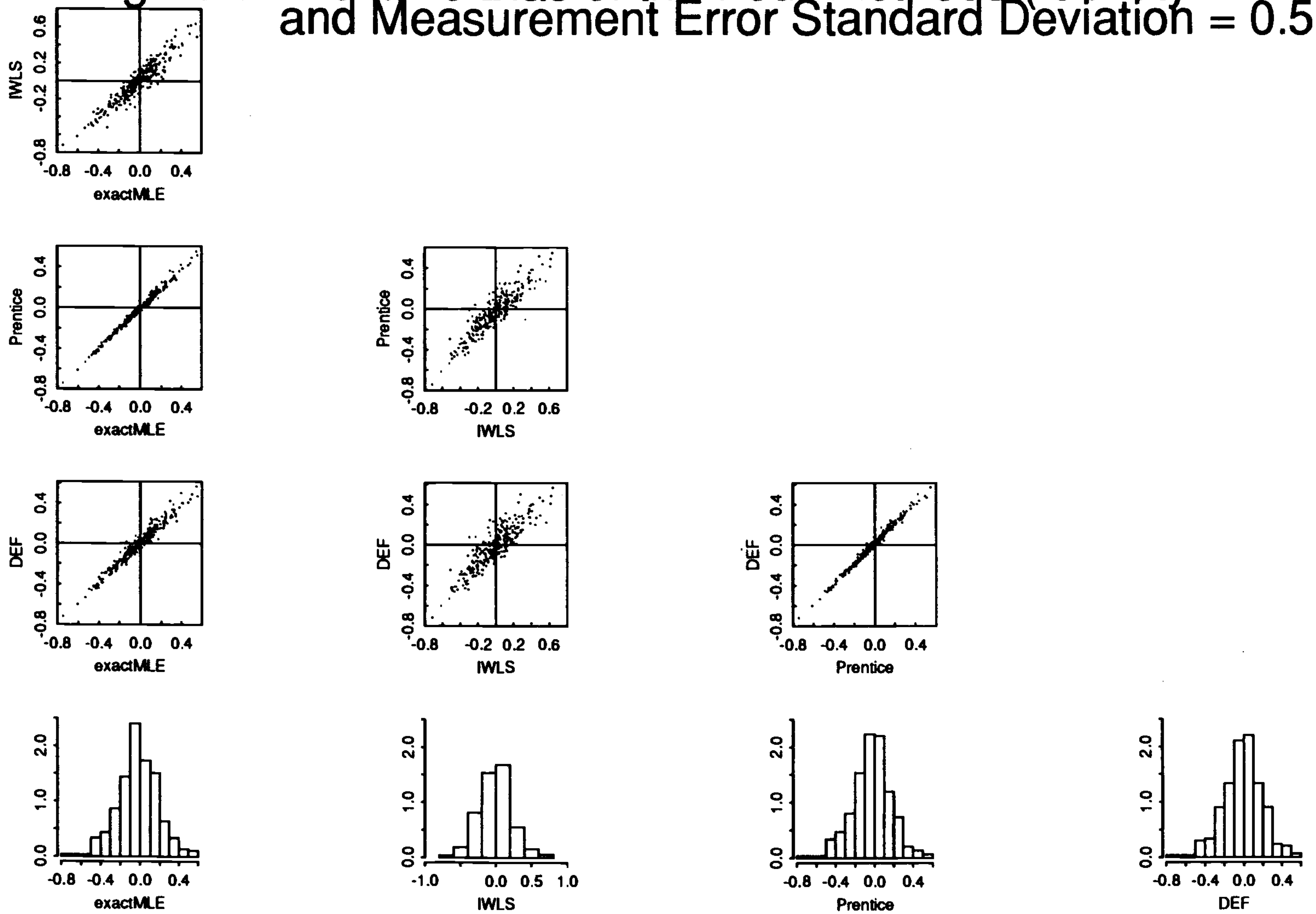


Figure 6. Relative Bias of the Four Methods (Identity Link and Measurement Error Standard Deviation = 0.5)



produce very similar estimates for all blocks (data sets) for small measurement error (.1), as is seen in figure 4. For larger multiplicative measurement error the differences are more pronounced. Figures 5 and 6 illustrate that Prentice's method produces estimates that are very highly correlated with the Lognormal MLE and that the DEF method is slightly less correlated. The IWLS method is the most variable in this simulation experiment but, as is seen in the histogram, is nearly as efficient as the other methods except for very large measurement errors.

Very similar conclusions can be drawn from the summary tables. Table 6 summarizes the relative bias for this situation (binomial with identity link). In general, the basic pattern of larger bias with larger measurement error held true. The naive estimator is considerably more biased than the other estimators; the bias is only about 2% when the standard deviation of d_i is .1 and about 44% when the standard deviation is .5. The simulation did not have enough

Table 6 --- Bias in $\hat{\beta}_1$ as a Proportion of β_1 (x 100)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	-2.26 (0.9)	-18.80 (0.9)	-43.85 (1.2)
Lognormal MLE	0.01 (0.9)	-0.06 (0.9)	-2.24 (1.1)
IWLS	0.17 (0.9)	0.02 (0.9)	-1.22 (1.1)
Prentice	0.08 (0.9)	-0.77 (0.9)	-2.99 (1.1)
DEF	0.14 (0.9)	0.01 (0.9)	-1.04 (1.1)

Standard error of estimates are in parentheses.

power to pick up differences in bias between the four estimators of interest although it is clear that, at least in comparison to the naive estimator, they are all essentially unbiased for these conditions.

The MSE increased with the measurement error for all estimators, see Table 7 for a summary of MSEs scaled by $1/\beta_1^2$. The MSE for the naive estimator is considerably larger than the other estimators except when the measurement error is small. The size of the relative MSEs follows the same pattern as the spread seen in the graphs. Thus, estimators can be ranked in terms of efficiency as Lognormal MLE, Prentice, DEF, and IWLS. However, for small and moderately sized measurement error there seems to be little practical difference between these methods.

Table 7 --- $MSE(\hat{\beta}_1)/\beta_1^2$ (x 100)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	2.48	6.20	22.63
Lognormal MLE	2.43	2.53	3.66
IWLS	2.43	2.57	3.95
Prentice	2.43	2.54	3.73
DEF	2.43	2.56	3.76

7.2.2 Binomial Model with the Logit Link Function

7.2.2.1 Design of the Simulation

This simulation study is designed to create data sets that fit into the situation described in section 5.5.1. The purpose is to compare the results for a non-identity link function using some of the same methods as the analysis in section 7.2.1. The set-up is the same as section 7.2.1, except that the identity link is replaced with the logit link in step 4). Thus step 4) becomes

4) $y_i \sim \text{bin}(100, p_i)/100$,

where $\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = 0.5 + 0.0075x_i$.

7.2.2.2 Results of the Simulation

Figure 7 contains pair-wise scatterplots of the 300 values of the four estimators, for a multiplicative measurement error standard deviation of .1. There is little difference between the estimators for this case. In figures 8 and 9, for measurement error standard deviations of .3 and .5 it is apparent that Prentice's and the DEF estimator are both highly correlated with the Lognormal MLE. The IWLS estimator is slightly more variable.

The basic pattern of larger relative bias with larger measurement error held true; see table 8 for a summary of the relative bias. For measurement error standard deviations of .3 and .5 the last three estimators contain more bias (as a proportion of β_1) than they did in the linear model. This may be due to the linearization approximation necessary to compute $E(x_i|z_i)$. The Lognormal MLE does not require such an approximation. At least, these biases are only about 3% for

Figure 7. Relative Bias of the Four Methods (Logit Link and Measurement Error Standard Deviation = 0.1)

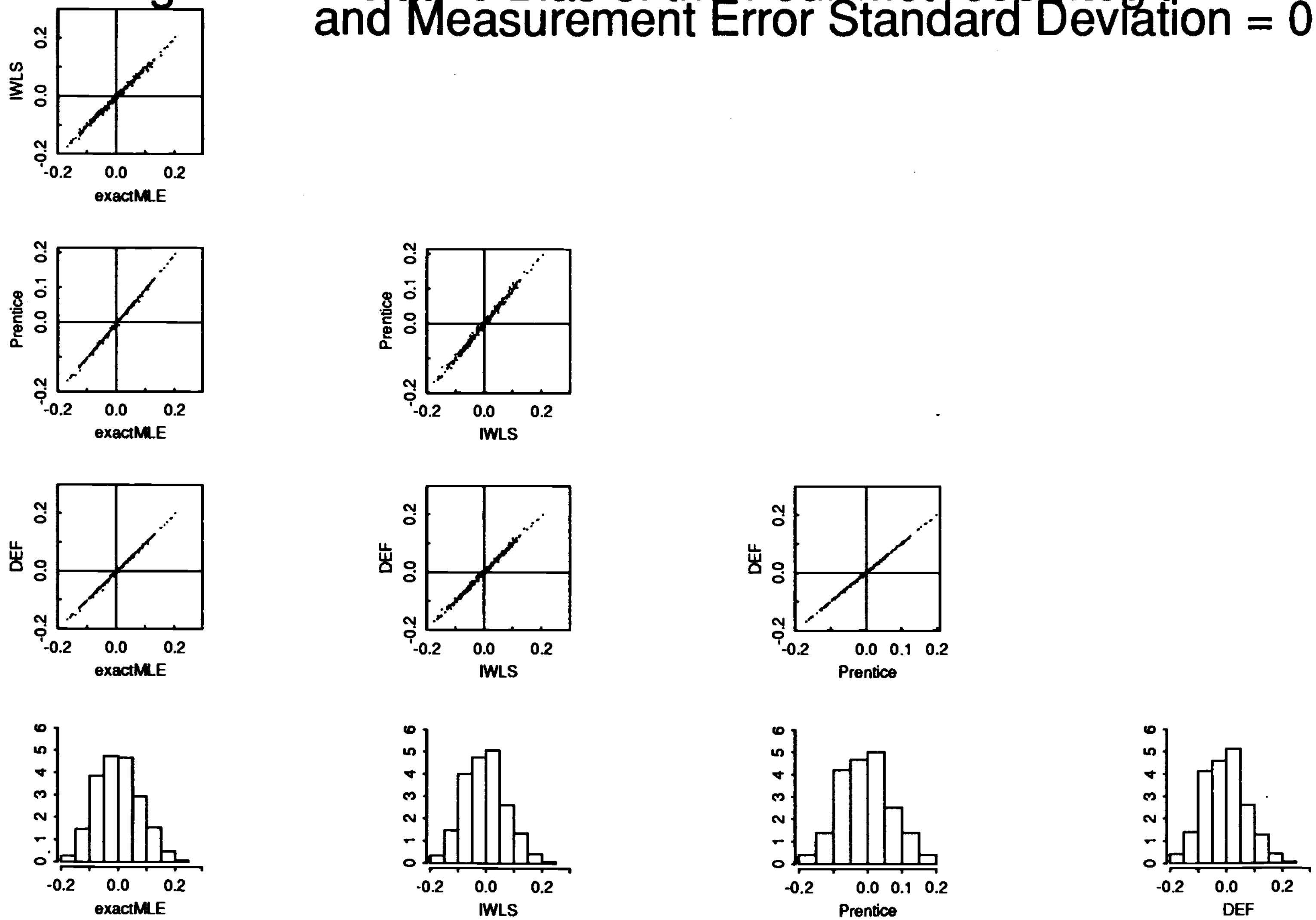


Figure 8. Relative Bias of the Four Methods (Logit Link and Measurement Error Standard Deviation = 0.3)

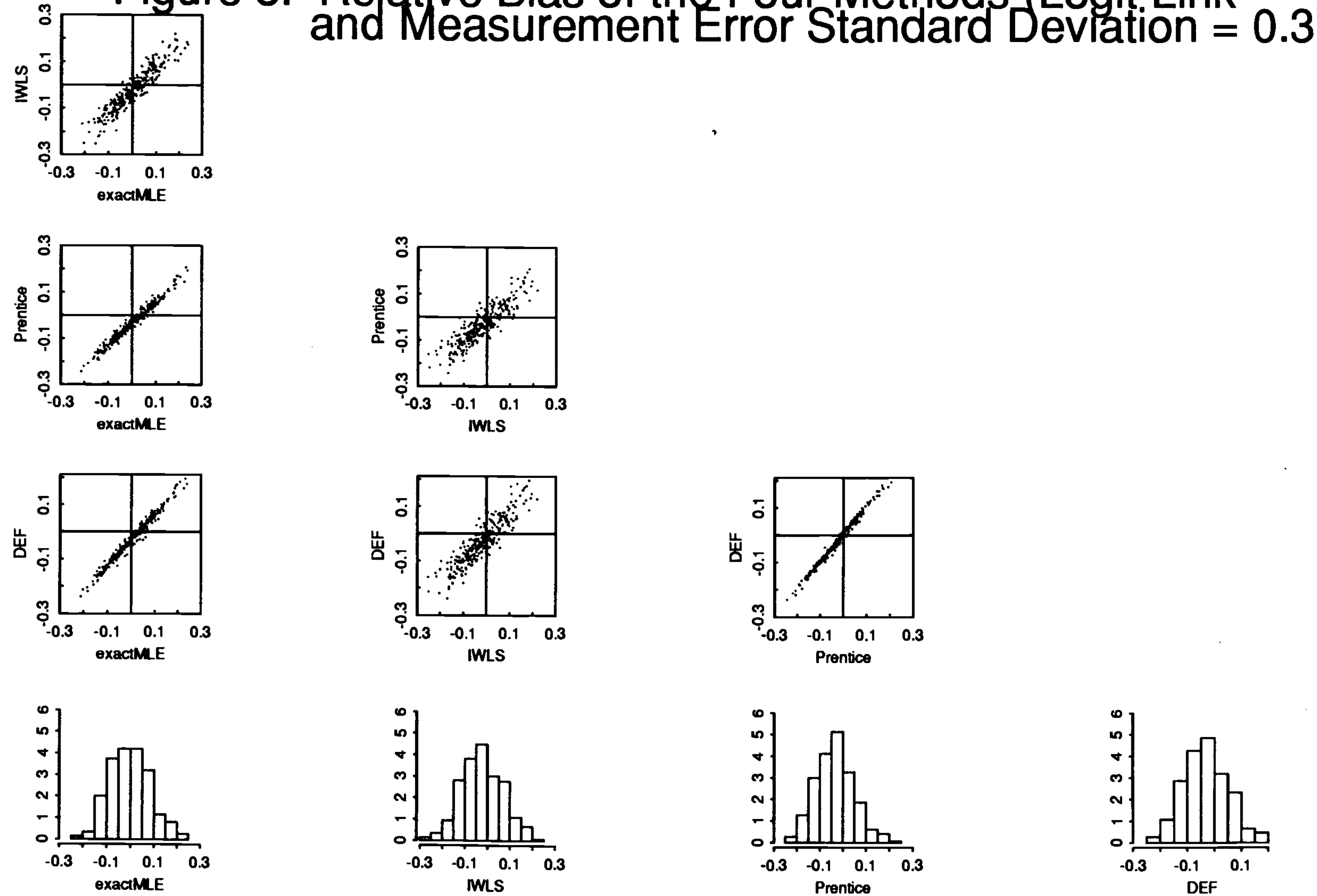
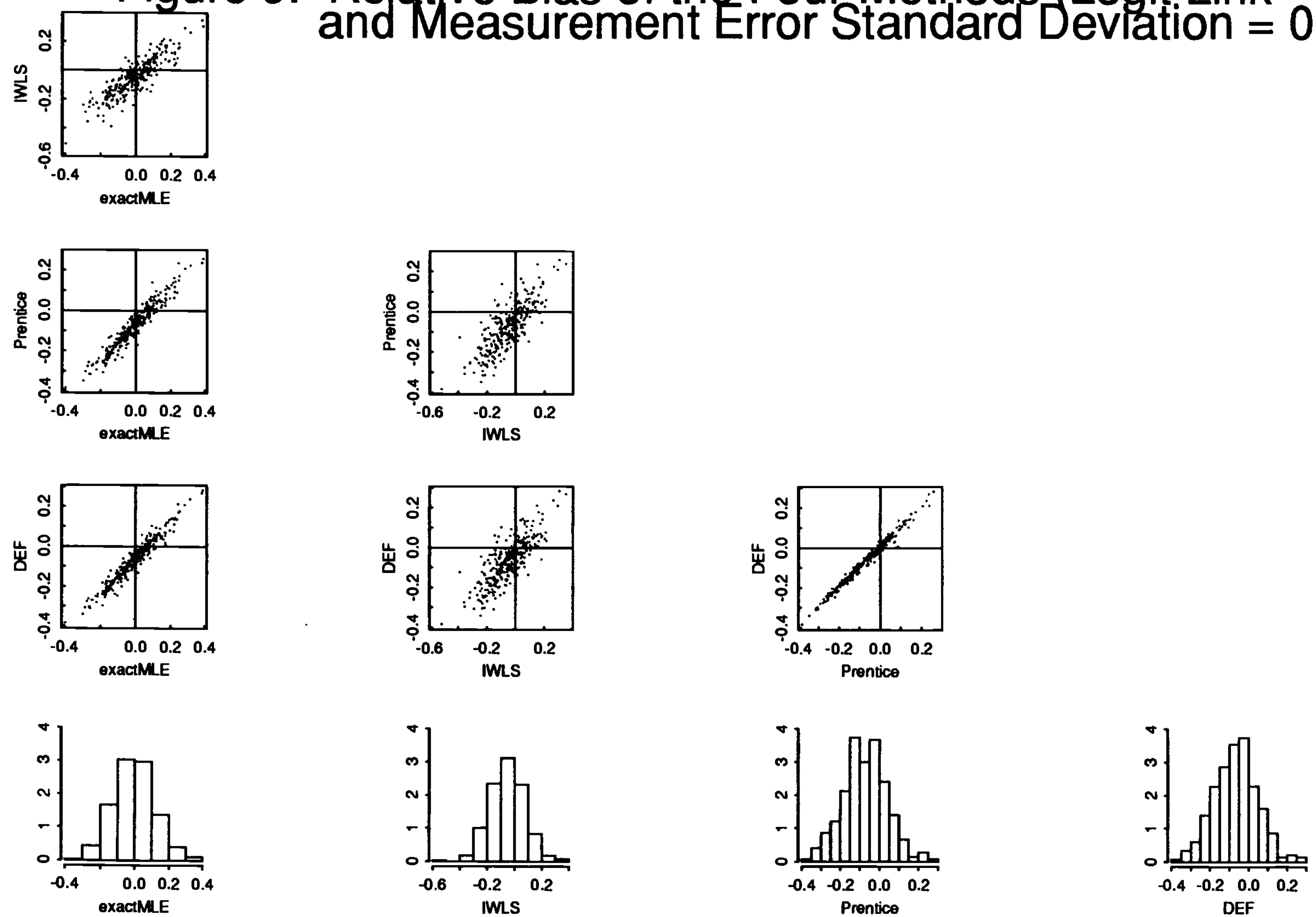


Figure 9. Relative Bias of the Four Methods (Logit Link and Measurement Error Standard Deviation = 0.5)



measurement errors with a standard deviation of .3. The larger the measurement error error, the wider the range of $g^{-1}(\cdot)$ we are trying to approximate with a straight line. IWLS, Prentice's, and the DEF methods appear to be giving very similar results, with IWLS doing slightly better especially in the presence of high measurement error.

Table 8 --- Bias in $\hat{\beta}_1$ as a Proportion of β_1 (x 100)
(Logit Link)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	-2.75 (0.4)	-22.10 (0.5)	-46.97 (0.7)
Lognormal MLE	0.01 (0.4)	-0.11 (0.5)	-0.58 (0.6)
IWLS	-0.36 (0.4)	-2.36 (0.5)	-4.96 (0.7)
Prentice	-0.49 (0.4)	-3.51 (0.5)	-7.23 (0.6)
DEF	-0.42 (0.4)	-3.11 (0.5)	-6.67 (0.6)

Standard error of estimate are in parentheses.

The MSE is larger for larger measurement errors as expected, see Table 9 for a summary of MSE scaled by $1/\beta_1^2$ for all methods. The naive estimator has considerably larger MSEs. All four of the remaining methods have similar MSEs for smaller values of measurement error, but for larger measurement error values Prentice and DEF have smaller MSE as is all seem in figures 8 and 9. Again, in comparison to the naive estimator, the others are quite similar. Based on tables 8 and 9 we can surmise that the Prentice and DEF methods are more affected by the linear approximation than is the IWLS method. The exact reason for this remains to be studied.

Table 9 --- $MSE(\hat{\beta}_1)/\beta_1^2$ (x 100)
(Logit Link)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	0.59	5.70	23.58
Lognormal MLE	0.51	0.68	1.17
IWLS	0.51	0.77	1.56
Prentice	0.52	0.81	1.73
DEF	0.51	0.78	1.64

7.3 Robustness

The issue of robustness is addressed using two situations. The simulation in Section 7.3.1 allows for the underlying model for the d_i's to be misspecified. In section 7.3.2, the underlying model for the distribution of the x_i's is incorrectly specified. Thus, in both situations the distribution of x_i given z_i assumed by the estimating procedures is incorrect. Two effects of these wrong assumptions are that the distribution of y_i given z_i used for the exact MLE is incorrectly specified and the mean and variance of y_i given z_i used in IWLS, Prentice's method and the DEF method are also incorrect. It would be better if, with the wrong underlying model, we could consider these effects separately, but they are confounded in the situations considered here.

7.3.1 Measurement Errors Distributed Contaminated Lognormal

7.3.1.1 Design of the Simulation

This simulation study is designed to create data sets that are binomial with identity link as described in section 7.2.1, except that the underlying model for the distribution of d_i is incorrect. The set-up is the same as section 7.2.1, except that the log d_i's in step 2) are no longer normally distributed. Step 2) becomes

2) log d_i ~ contaminated normal, where there is a 90% chance of log d_i being normal with mean 0 and variance σ_d^2 and a 10% chance of log d_i being normal with mean 0 and variance $9\sigma_d^2$, the resulting mean and variance of the log d_i's is $(0,\ 1.8\sigma_d^2)$.

The data is analyzed as if it is generated as described in section 7.2.1.1, where the log d_i are normal.

7.3.1.2 Results of the Simulation

Figure 10 compares the four methods for a measurement error standard deviation of .1. All methods seem to be giving very similar estimates of slope for this value of measurement error. The relative bias is bigger as evident by the larger horizontal scale of the plots. Figures 11 and 12 exhibit more variability as is seen in earlier situations. The comparative results for the different methods look very similar to the results from section 7.2.1.2.

As is true for section 7.2, very similar results can be concluded from the summary tables. The basic pattern of larger relative bias with larger measurement error held true; see table 10 for a summary of the relative bias. Relative bias does seem to be affected some by the

Figure 10. Relative Bias of the Four Methods (log d's Contaminated Normal and Measurement Error Standard Deviation = 0.1)

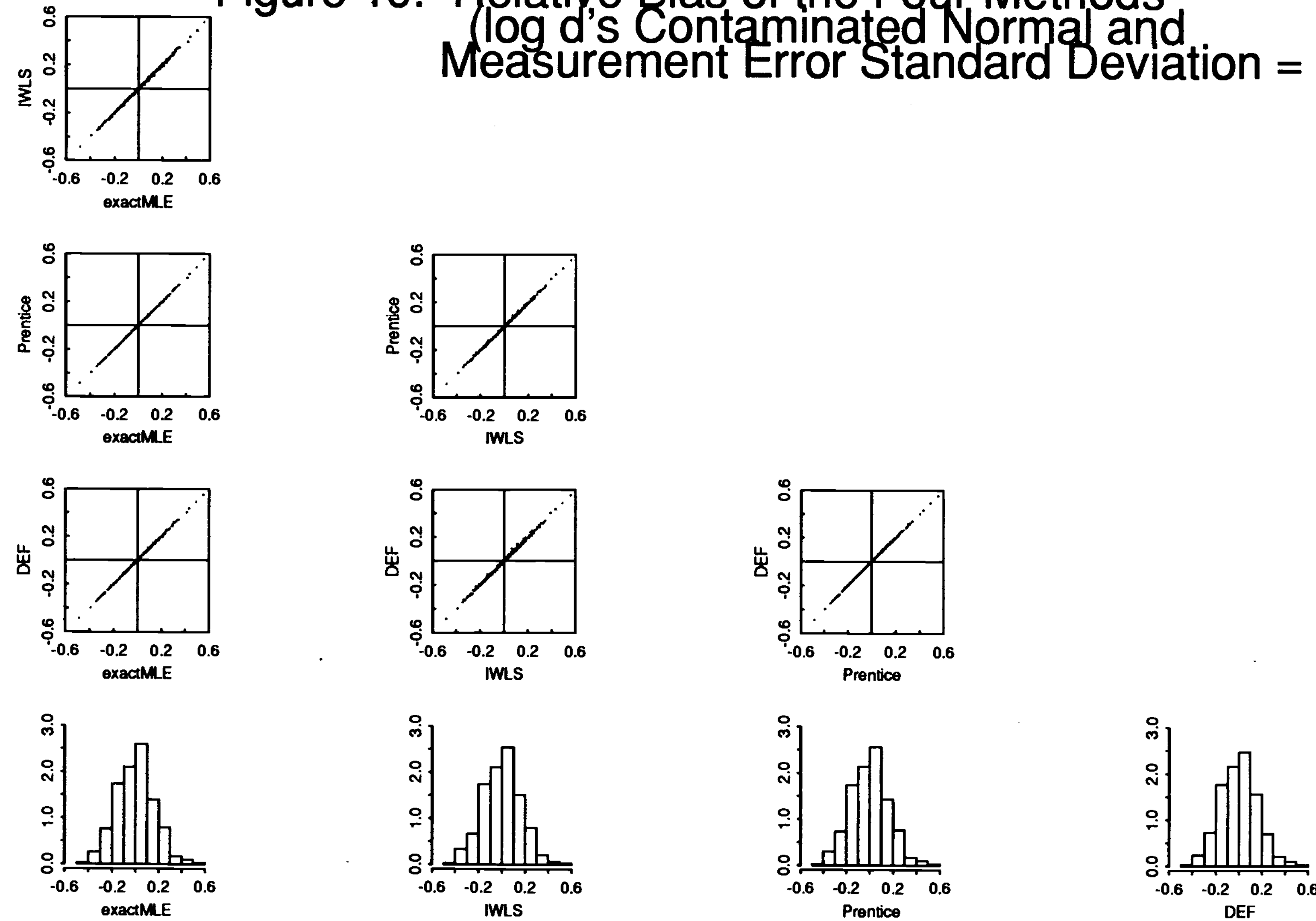


Figure 11. Relative Bias of the Four Methods
(log d's Contaminated Normal and
Measurement Error Standard Deviation = 0.3)

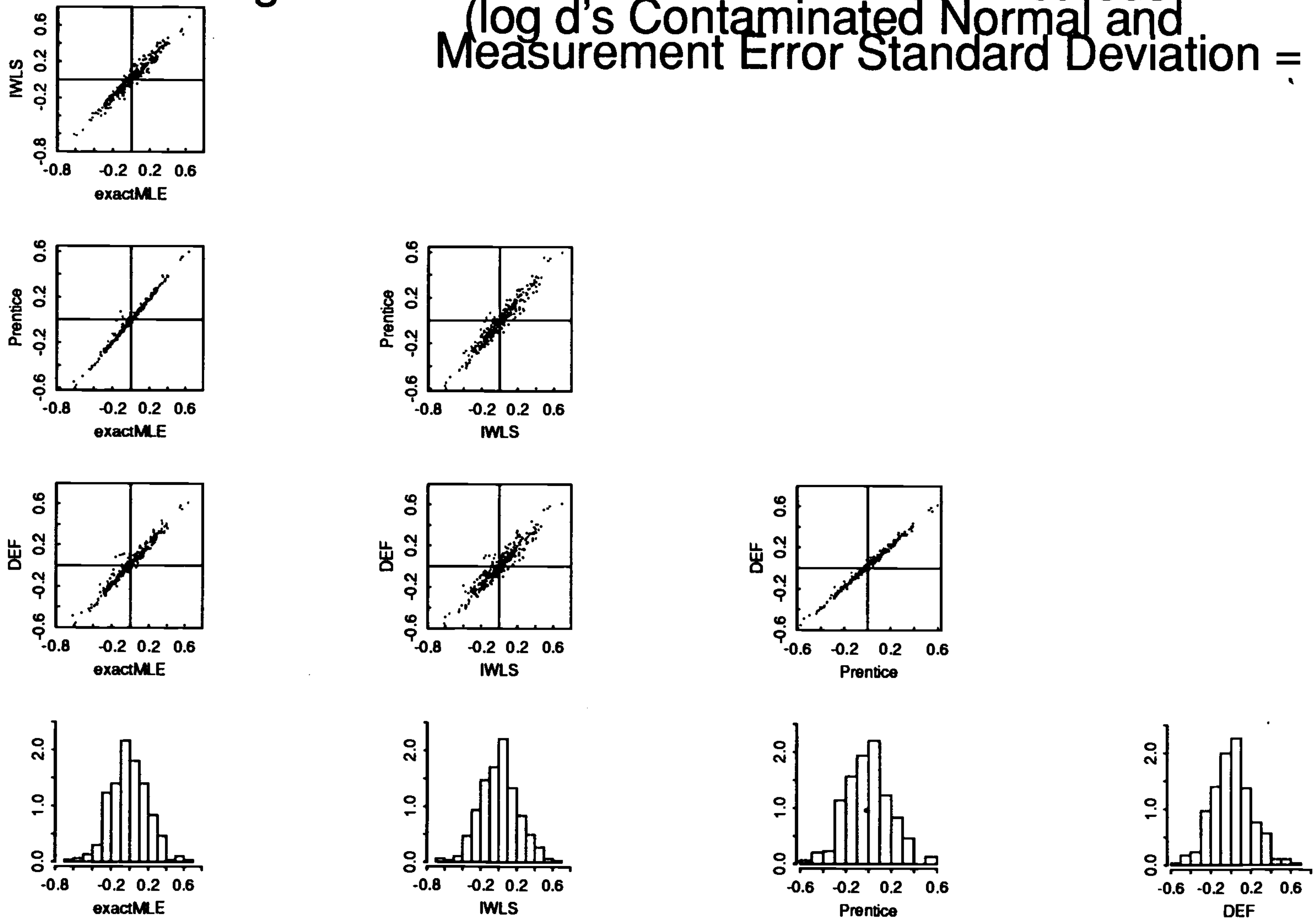
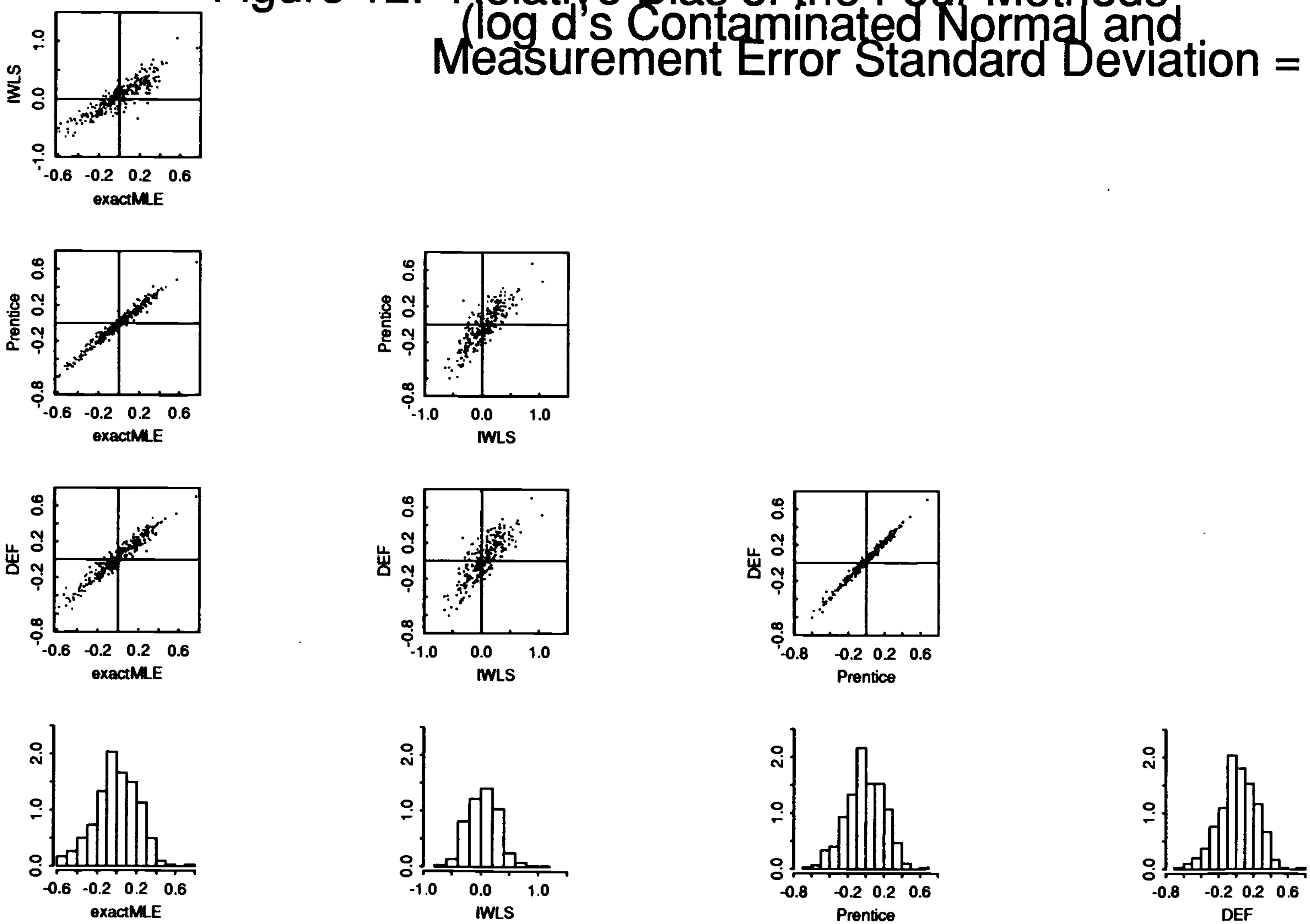


Figure 12. Relative Bias of the Four Methods (log d's Contaminated Normal and Measurement Error Standard Deviation = 0.5)



incorrect assumption about the distribution of log d_i. The relative MSE (table 11) does seem to be larger than when the distribution of the log d_i's is correctly assumed in section 7.2.

Table 10 --- Bias in $\hat{\beta}_1$ as a Proportion of β_1 (x 100)
(Identity Link and log d's Contaminated Normal)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	-4.20 (1.1)	-32.88 (1.2)	-61.68 (1.4)
Lognormal MLE	-0.11 (0.9)	-0.99 (1.1)	-0.22 (1.2)
IWLS	0.04 (0.9)	0.38 (1.1)	4.66 (1.3)
Prentice	-0.07 (0.9)	-0.81 (1.1)	-1.12 (1.2)
DEF	0.09 (0.9)	0.97 (1.1)	2.19 (1.2)

Standard error of estimate are in parentheses.

Table 11 --- $MSE(\hat{\beta}_1)/\beta_1^2$ (x 100)
(Identity Link and log d's Contaminated Normal)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	2.84	15.23	43.93
Lognormal MLE	2.65	3.71	4.12
IWLS	2.65	3.80	5.15
Prentice	2.65	3.76	4.30
DEF	2.65	3.85	4.46

7.3.2 Explanatory Variable Distributed Weibull

7.3.2.1 Design of the Simulation

This simulation study is designed to create data sets that are binomial with identity link as described in section 7.2.1, except that the underlying model for the distribution of x_i is different from what is assumed. The set-up is the same as section 7.2.1, except that the log x_i's in step 1) are no longer normally distributed. Step 1) becomes

1) $x_i \sim$ Weibull(1.25, 77.5), (this has approximately the same mean and variance as the x_i in section 7.2.1.

The data is analyzed as if it is generated as described in section 7.2.1.1, where the log x_i are normal.

7.3.2.2 Results of the Simulation

Figure 13 compares the four methods for a measurement error standard deviation of .1. All methods seem to be giving very similar estimates of slope for this value of measurement error. The relative bias is bigger as evident by the larger scales of the plots. Figures 14 and 15 have similar variability compared to section 7.2.1. The comparative results for the different methods look very similar to the results from section 7.2.1.2.

The basic pattern of larger relative bias with larger measurement error held true; see table 12 for a summary of the relative bias. Relative bias does seem to be larger when the assumption about the distribution of x_i is incorrect. The summary of MSEs is presented in table 13.

Figure 13. Relative Bias of the Four Methods (x's Weibull and Measurement Error Standard Deviation = 0.1)

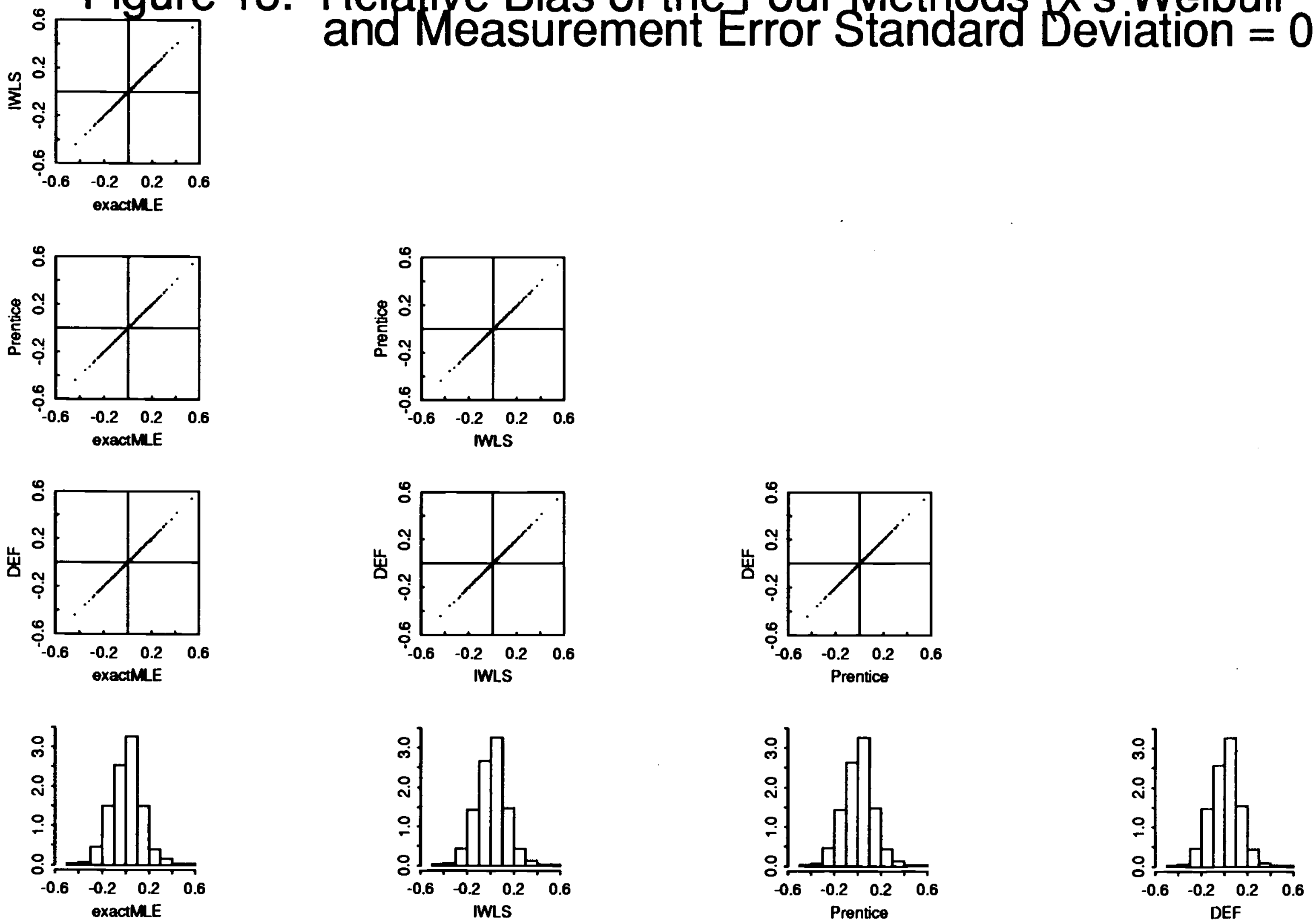


Figure 14. Relative Bias of the Four Methods (x's Weibull and Measurement Error Standard Deviation = 0.3)

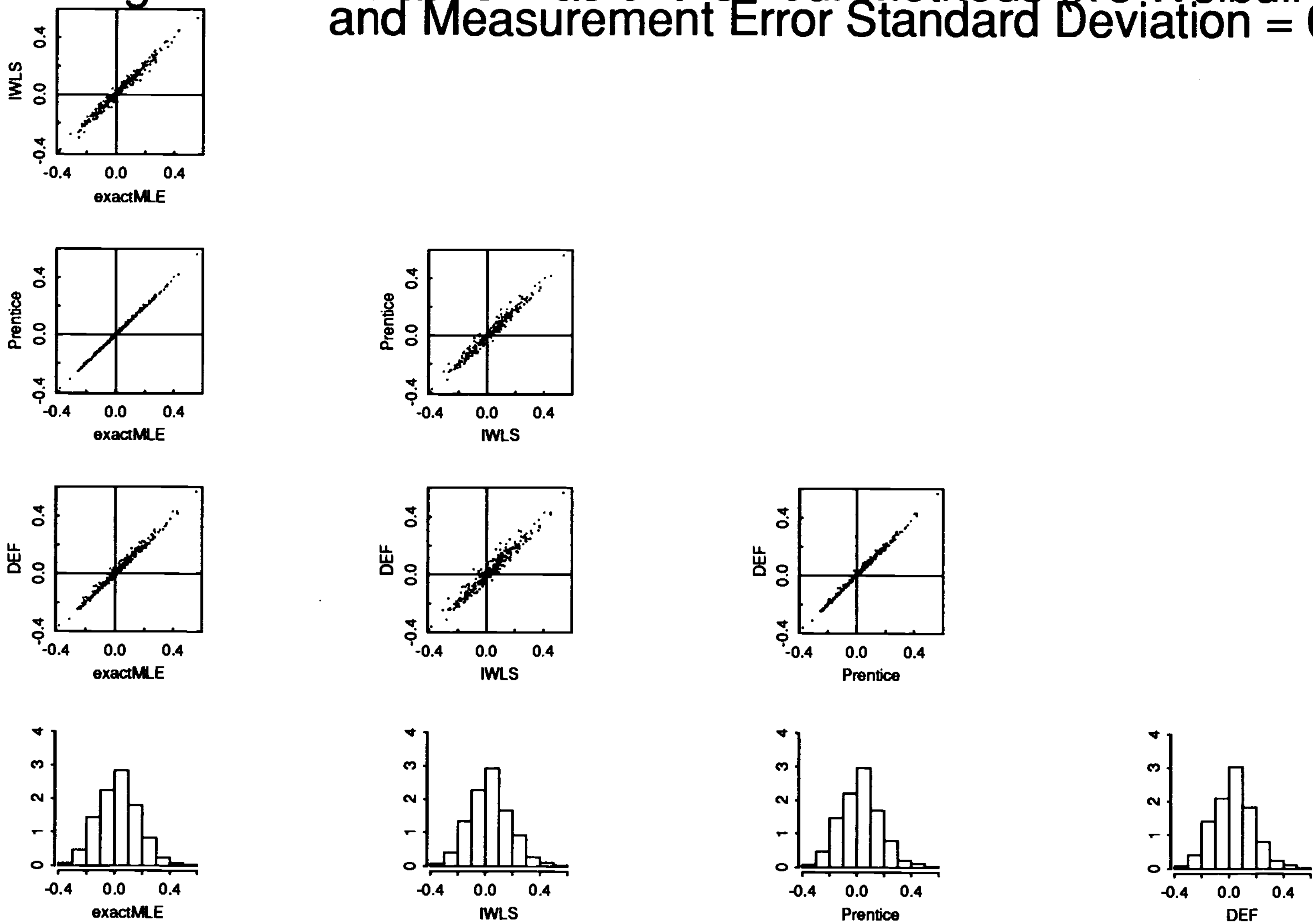


Figure 15. Relative Bias of the Four Methods (x's Weibull and Measurement Error Standard Deviation = 0.5)

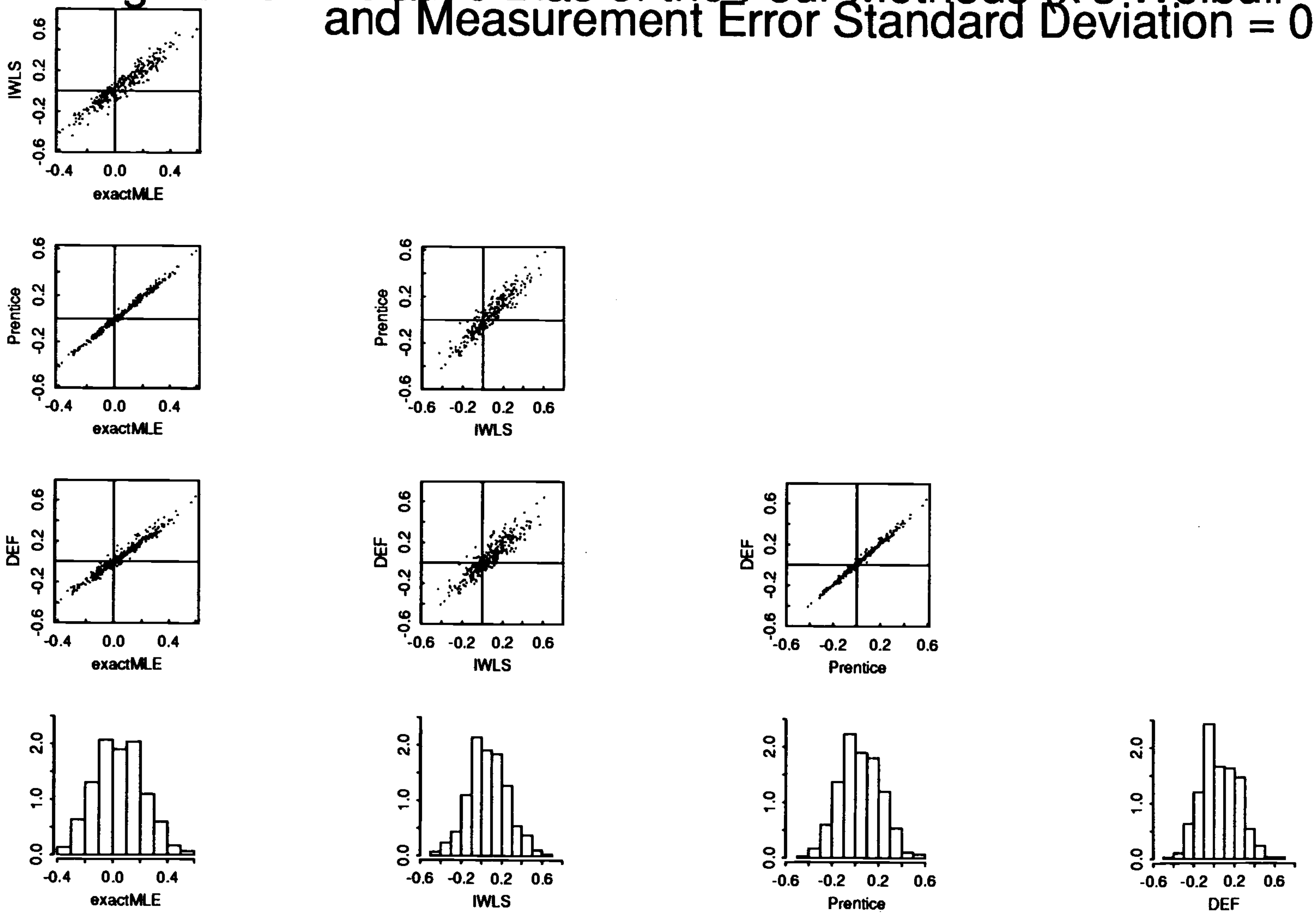


Table 12 --- Bias in $\hat{\beta}_1$ as a Proportion of β_1 (x 100)
(Identity Link and x's Weibull)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	-1.35 (0.8)	-15.20 (0.8)	-39.11 (1.1)
Lognormal MLE	0.74 (0.8)	2.90 (0.8)	4.96 (1.0)
IWLS	0.80 (0.8)	3.38 (0.8)	6.30 (1.0)
Prentice	0.72 (0.8)	2.73 (0.8)	3.98 (1.0)
DEF	0.75 (0.8)	3.26 (0.8)	5.18 (1.0)

Standard error of estimate are in parentheses.

Table 13 --- $MSE(\hat{\beta}_1)/\beta_1^2$ (x 100)
(Identity Link and x's Weibull)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	1.76	4.43	18.61
Lognormal MLE	1.75	2.08	2.97
IWLS	1.75	2.14	3.29
Prentice	1.75	2.08	2.91
DEF	1.75	2.14	3.09

7.4 Sample Estimates Used to Estimate Nuisance Parameters

7.4.1 Design of the Simulation

The parameters μ_x, σ_x^2, and σ_d^2 are nuisance parameters in this problem. Because this is a simulation these values are known, but in practice it is likely that they would have to be estimated from knowledge about σ_d^2 and sample statistics of the z_i's (see Carroll et al, 1984). Since the estimation of β_1 is dependent on the nuisance parameters, it is desired to know how using estimates rather than the true values affects the estimation of β_1. For this case the data is generated as in section 7.2.1, but analyzed using the sample estimates for the parameters μ_x, σ_x^2, and σ_d^2, i.e. if $z_i^* = \log z_i$, then

$$\hat{\mu}_x = \bar{z}_i^*,$$

$$\sigma_x^2 = s_z^2 - \sigma_d^2, \text{ and}$$

$$R = \frac{s_z^2 - \sigma_d^2}{s_z^2}.$$

7.4.2 Results of the Simulation

Figures 16, 17 and 18 contain pair-wise scatterplots of the estimates based on 300 simulated samples for the four methods and the three measurement error standard deviations. The plots look very similar to those in section 7.2.1.2, although there appears to be slightly more bias and bigger relative MSEs for the largest measurement error standard deviations. These facts also are supported by the summary tables 14 and 15. The estimation process does not appear to suffer greatly when sample estimates are used for these nuisance parameters.

Figure 16. Relative Bias of the Four Methods (Sample Estimates Used and Measurement Error Standard Deviation = 0.1)

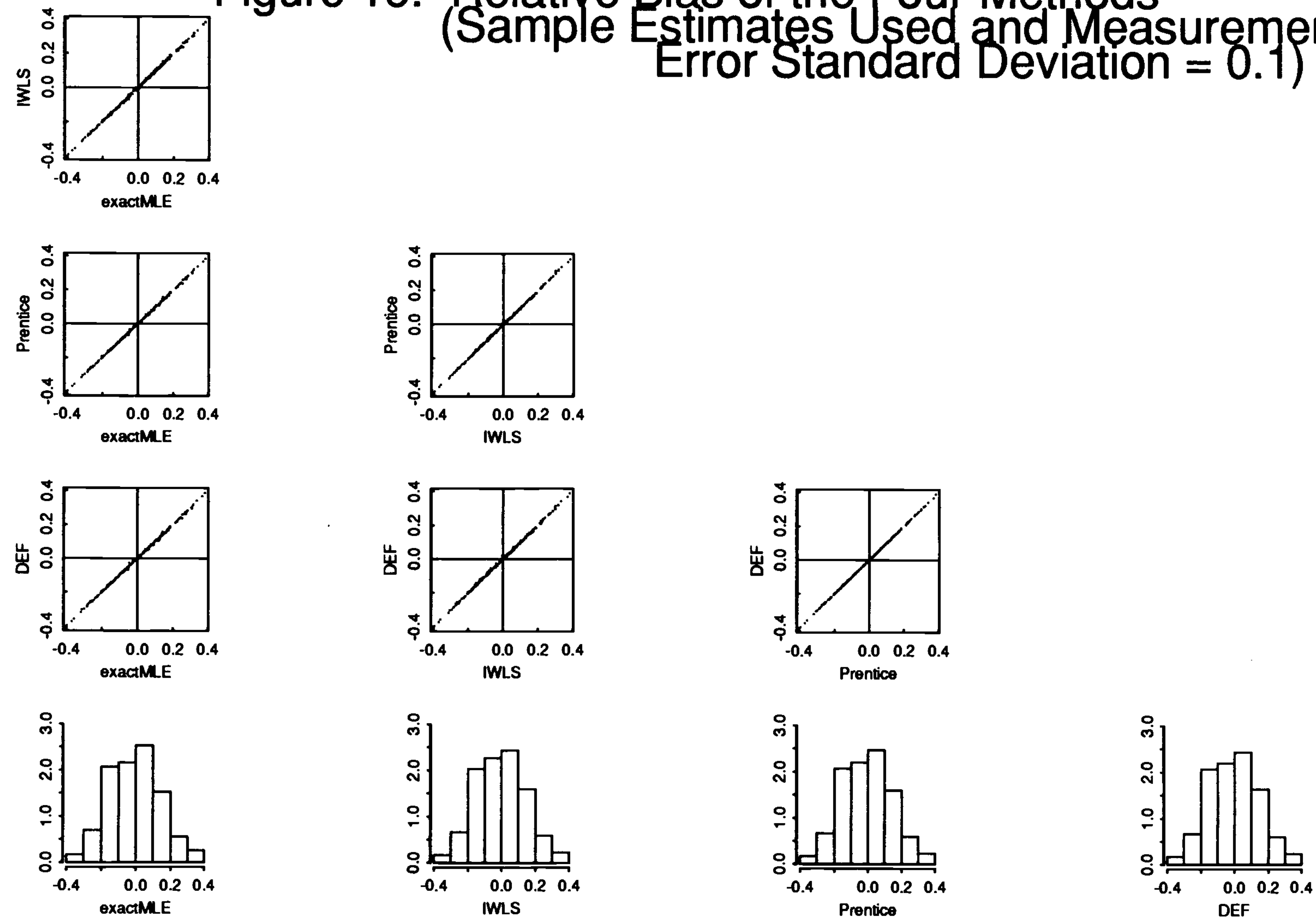


Figure 17. Relative Bias of the Four Methods (Sample Estimates Used and Measurement Error Standard Deviation = 0.3)

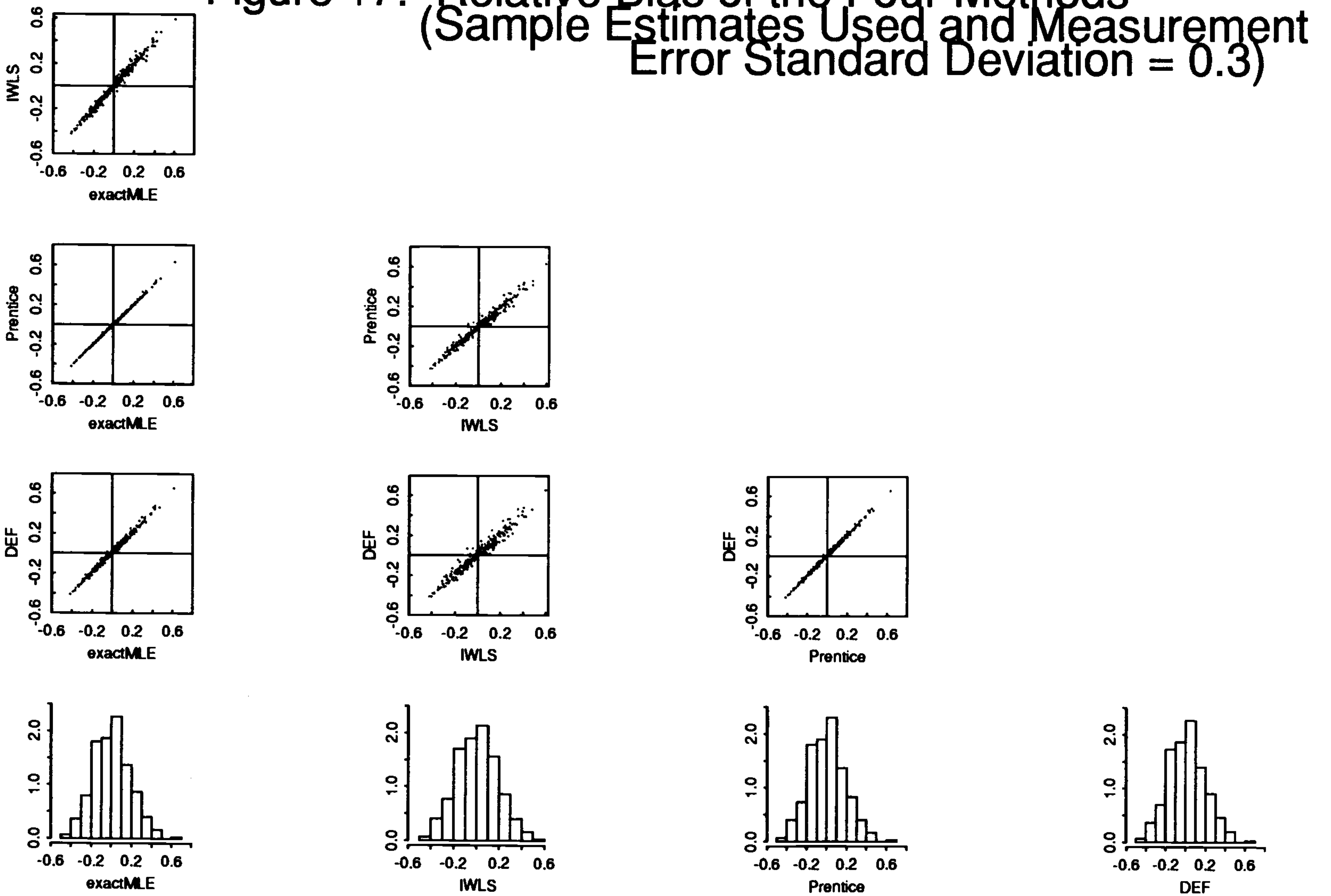


Figure 18. Relative Bias of the Four Methods (Sample Estimates Used and Measurement Error Standard Deviation = 0.5)

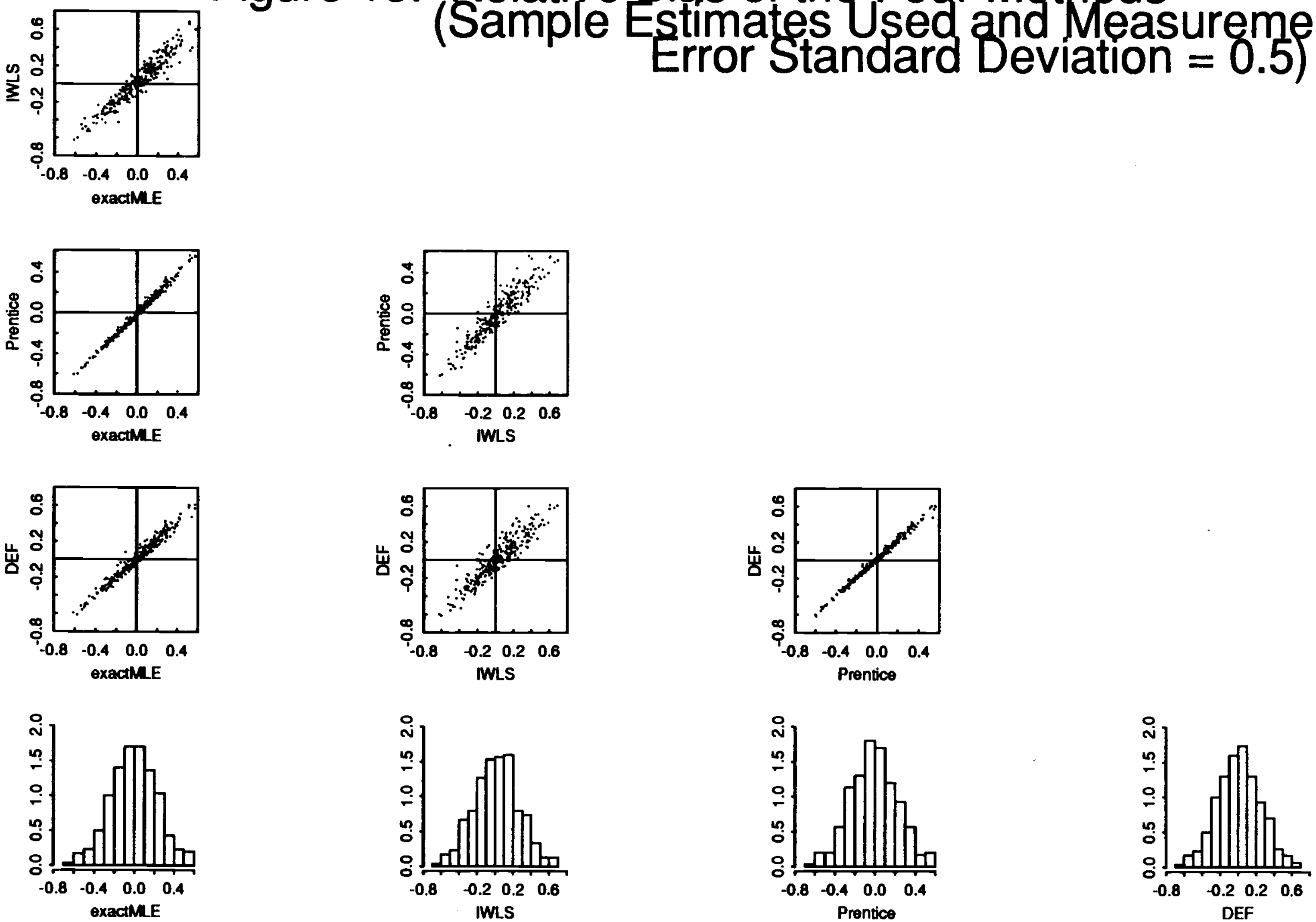


Table 14 --- Bias in $\hat{\beta}_1$ as a Proportion of β_1 (x 100)
(Identity Link and Sample Estimates Used)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	-3.06 (0.9)	-19.15 (1.1)	-43.13 (1.4)
Lognormal MLE	-0.81 (0.9)	0.30 (1.0)	-0.11 (1.2)
IWLS	-0.71 (0.9)	0.71 (1.0)	1.72 (1.3)
Prentice	-0.77 (0.9)	0.27 (1.0)	-0.97 (1.2)
DEF	-0.70 (0.9)	1.28 (1.0)	1.05 (1.2)

Standard error of estimate are in parentheses.

Table 15 --- $MSE(\hat{\beta}_1)/\beta_1^2$ (x 100)
(Identity Link and Sample Estimates Used)

Method	Measurement Error Standard Deviation		
	0.1	0.3	0.5
Naive	2.28	4.98	24.17
Lognormal MLE	2.19	3.08	4.46
IWLS	2.19	3.12	4.80
Prentice	2.19	3.09	4.51
DEF	2.19	3.14	4.65

7.5 Standard Error Used to Estimate the Standard Deviation of the Estimate

Figures 19 through 23 summarize the comparison between the (Monte Carlo) standard deviation of the estimate displayed here as a proportion of β_1 (SD/β_1) and the average of the standard errors displayed here as SE/β_1 calculated from the asymptotic formulas for each sample. For these comparisons 1000 data sets were generated for each combination of the following situations with each of the measurement error standard deviations of (0.2, 0.3, 0.4, 0.5):

Situation 1 -- y_i is binomial, log x_i is normal, log d_i is normal and the identity link used as described in section 7.2.1. The data for each of these four measurement error standard deviations is analyzed assuming the correct distributions for y_i, x_i, and d_i. The true values for the nuisance parameters (μ_x, σ_x^2, R) are used.

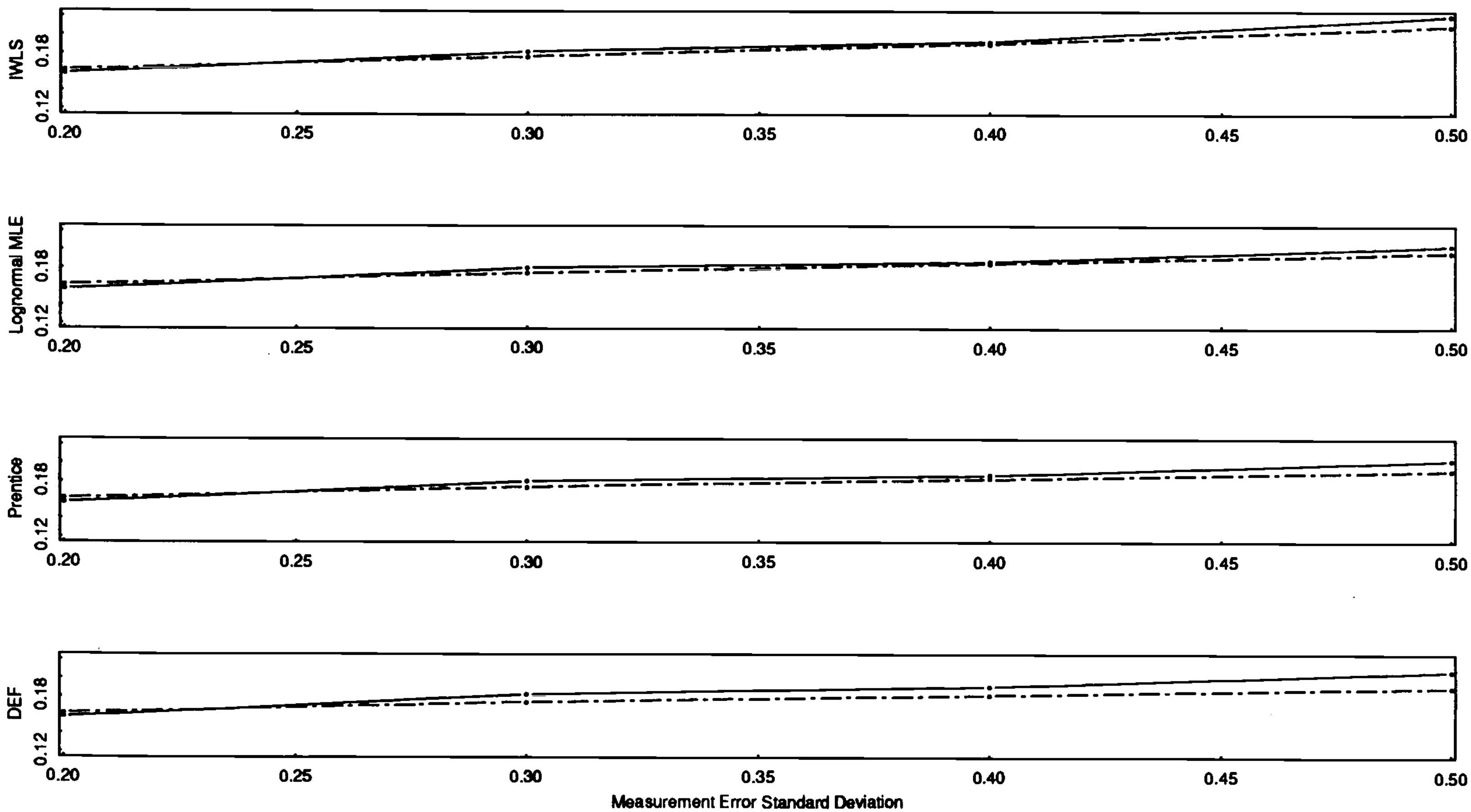
Situation 2 -- Same as situation 1, except the logit link is used. This situation is described in section 7.2.2. This data is analyzed with the correct underlying model used.

Situation 3 -- Same as situation 1, except the log d_i's are contaminated normal. This data is analyzed as in situation 1. Thus, the wrong underlying model for the d_i's is used. This situation is described in section 7.3.1.

Situation 4 -- Same as situation 1, except the x_i's are Weibull. This data is analyzed as in situation 1. Thus, the wrong underlying model for the x_i's is used. This situation is described in section 7.3.2.

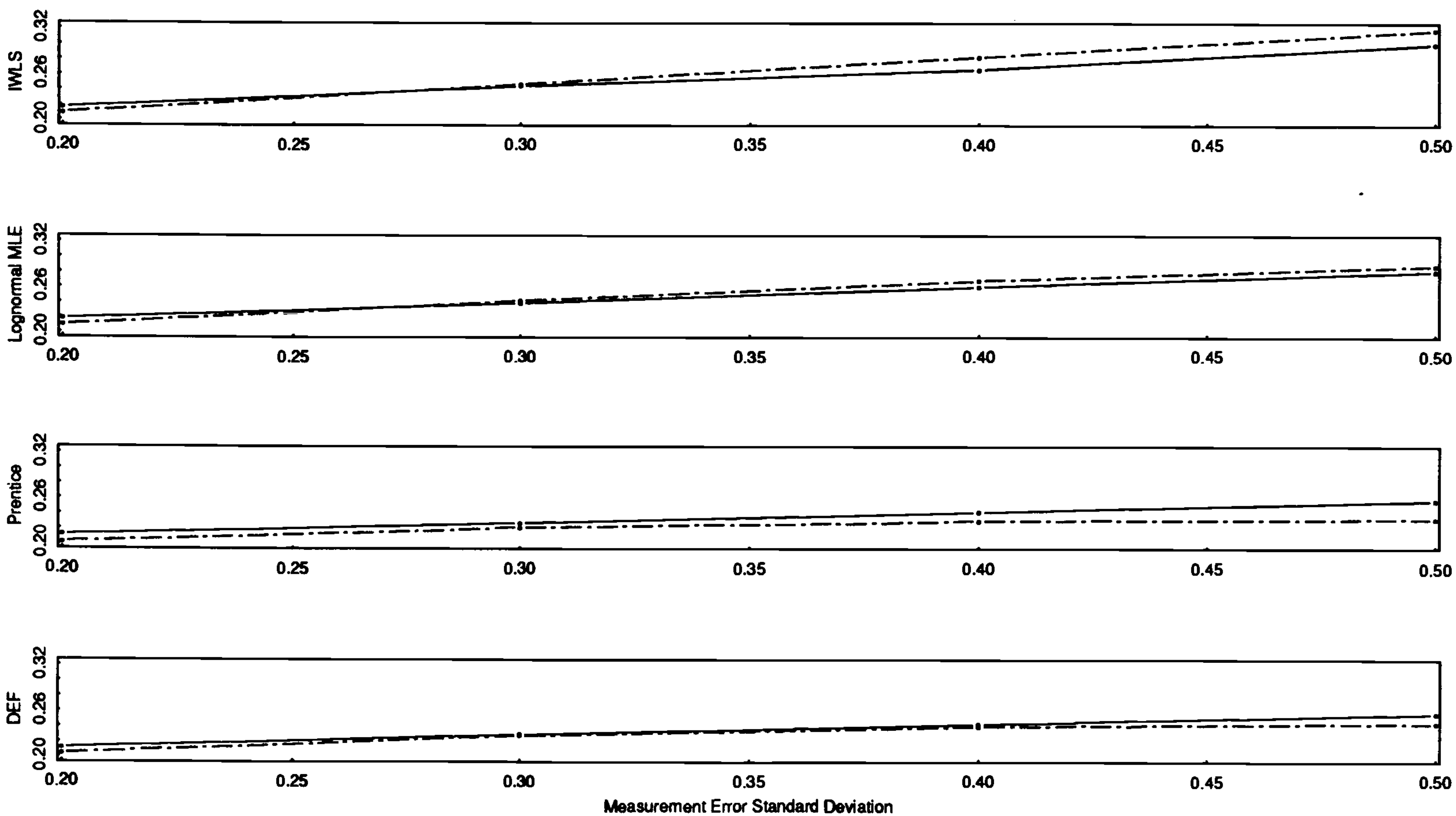
Situation 5 -- Same as situation 1, except sample estimates of the

Figure 19. SD of Estimate vs. Average SE (Identity Link)



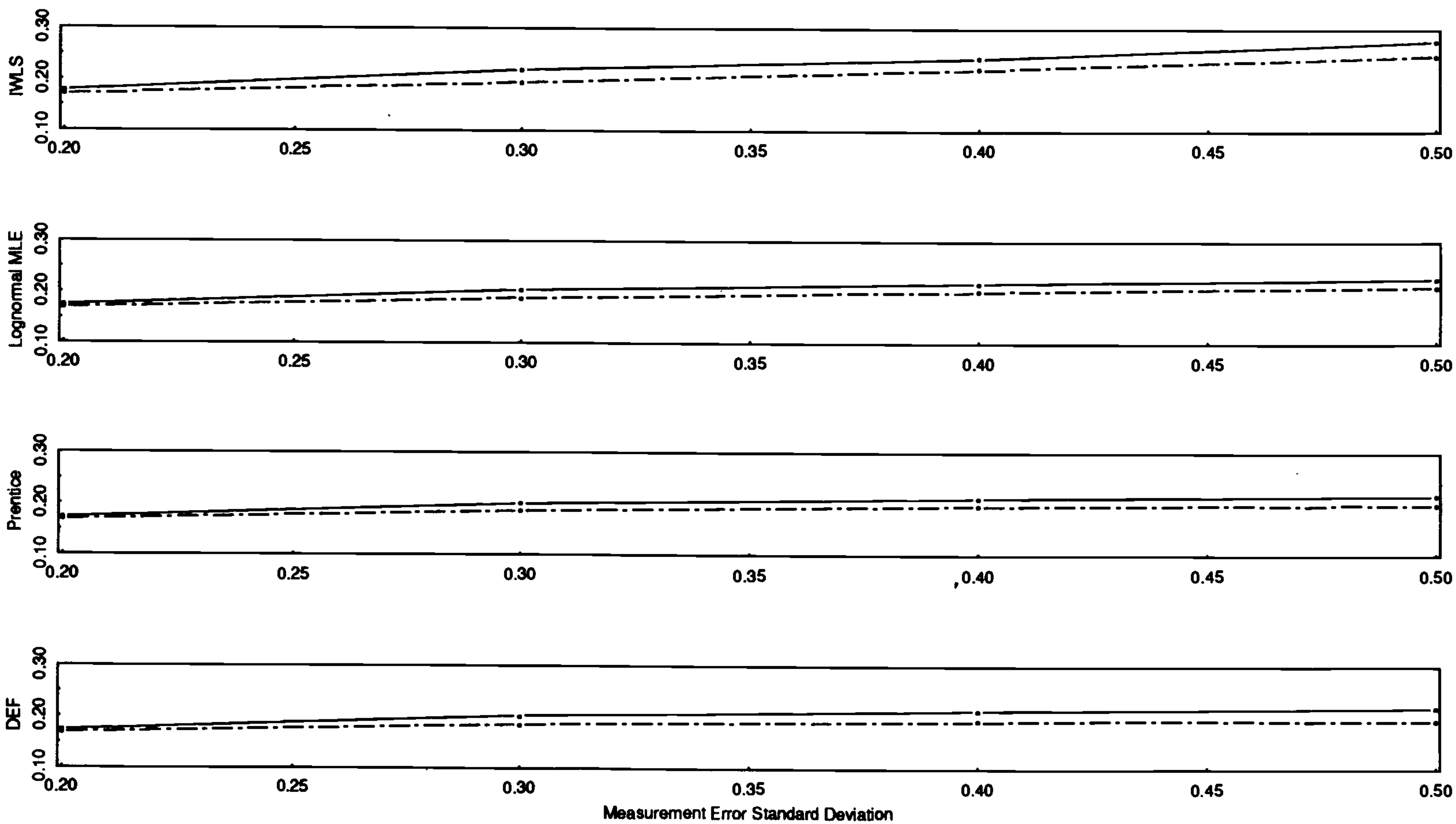
SD of Estimate (as a Proportion of Slope) -------- Average SE (as a Proportion of Slope) - - - -

Figure 20. SD of Estimate vs. Average SE (Logit Link)



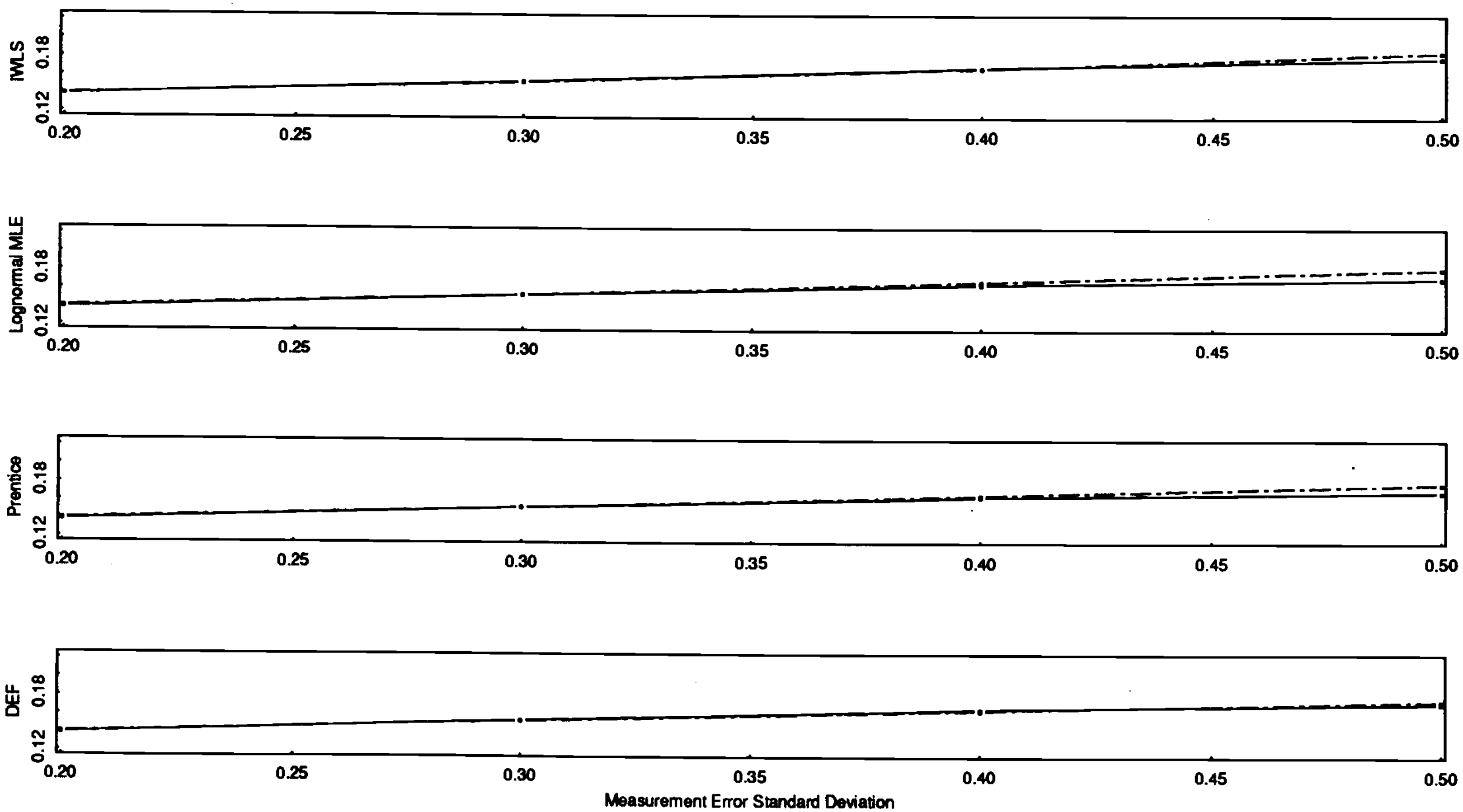
SD of Estimate (as a Proportion of Slope) -------- Average SE (as a Proportion of Slope) - - - -

Figure 21. SD of Estimate vs. Average SE (Log d Contam. Normal)



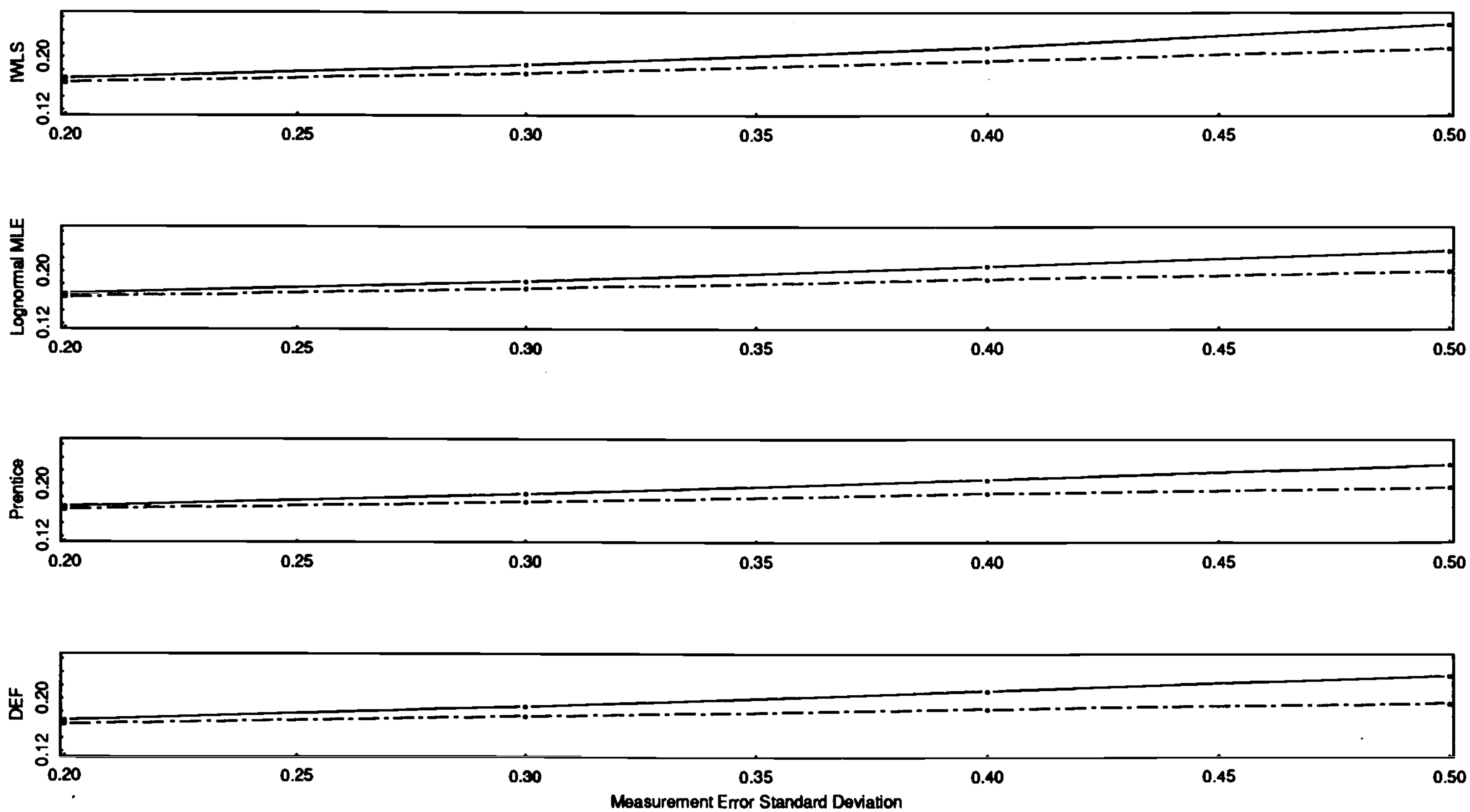
SD of Estimate (as a Proportion of Slope) ------- Average SE (as a Proportion of Slope) - - - -

Figure 22. SD of Estimate vs. Average SE (X is Weibull)



SD of Estimate (as a Proportion of Slope) -------- Average SE (as a Proportion of Slope) - - - -

Figure 23. SD of Estimate vs. Average SE (Sample Estimates Used)



SD of Estimate (as a Proportion of Slope) -------- Average SE (as a Proportion of Slope) - - - -

nuisance parameters are used. This data is analyzed as in situation 1. This situation is described in section 7.4.

The general pattern is that the SE's underestimate the standard deviation of the estimate. The four methods of interest, IWLS, Lognormal MLE, Prentice, and DEF, appear to be doing an adequate job of estimating the standard deviation of the estimate for small and moderate measurement errors. For larger measurement error standard deviations the difference is more noticeable in most cases.

The standard deviation of the estimates and the standard errors do not seem to be affected much by the use of sample estimates of the nuisance parameters, see figure 23. Whereas the differences between the standard deviation of the estimates and the standard errors are larger in situation 3, figure 21. It seems surprising that the violation in situation 4 (figure 22) resulted in the variability of the estimates being reduced, as is also true for relative MSE. The standard deviation of the estimates for the DEF method are just slightly larger than the Lognormal MLE in some situations, but Prentice's method does better at producing similar results to the Lognormal MLE for the identity link situations. The DEF method does slightly better for the situation with the logit link.

Chapter 8

Conclusions

8.1 Summary of Double Exponential Family Maximum Likelihood Method

This thesis presents the double exponential family model as a means for estimating linear parameters of generalized linear models when explanatory variables contain random measurement error. The use of the procedure on several cases of interest has been illustrated and the method has been compared to several other estimators with similar purposes and to the exact maximum likelihood estimator (which requires numerical integration) for a few special cases.

The mean and variance in the generalized linear regression may be written as

$$E(y_i|x_i) = \mu_i = g^{-1}(\beta_0+\beta_1 x_i) \text{ and}$$

$$\mathrm{Var}(y_i|x_i) = V(\mu_i).$$

If z_i is the measurement of x_i then the density function for y_i given z_i is

$$\int f(y_i|x_i)f(x_i|z_i)dx_i, \tag{8.1}$$

which is usually quite difficult to use for estimation. On the other hand, it is easy to see that the mean and variance for y_i given z_i are

$$E(y_i|z_i) = \mu_i^* = E[g^{-1}(\beta_0+\beta_1 x_i)|z_i] \doteq g^{-1}[\beta_0+\beta_1 E(x_i|z_i)], \text{ and}$$

$$\mathrm{Var}(y_i|z_i) \doteq E[V(\mu_i)|z_i] + \{g^{-1}[\beta_0+\beta_1 E(x_i|z_i)]\}^2\beta_1^2\mathrm{Var}(x_i|z_i).$$

A double exponential family density is similar to the corresponding one parameter exponential family, except for the addition of an extra parameter to model the dispersion. The proposed method supposes that the density of y_i given z_i is a double exponential family density with this mean and variance and uses maximum likelihood techniques to estimate β.

8.2 Exact Maximum Likelihood, Approximate Maximum Likelihood and Iteratively Weighted Least Squares Methods

The methods that have been compared may be listed in order of sophistication in the following way:

1) The exact maximum likelihood estimator based on (8.1).

2) The approximate maximum likelihood estimators based on an approximation to (8.1)
   - i) with the beta-binomial distribution used as the approximation and
   - ii) with a double exponential family distribution used as the approximation.

3) The iteratively weighted least squares (IWLS) estimator based on $E(y_i|z_i)$ and $\mathrm{Var}(y_i|z_i)$.

The exact distribution of y_i given z_i as given in (8.1) does not have a closed-form solution for the situations that are simulated in chapter 7. Thus, the estimation of β involves numerical integration.

The conceptual idea of maximum likelihood is straight forward, but is complicated by not having a closed form solution.

The idea of the approximations in 2) above is to arrive at an approximation to (8.1) that has solutions close to the exact maximum likelihood estimator but does not involve numerical integration. Prentice (1986) uses the beta-binomial distribution to approximate (8.1). He comments in his paper that the use of the beta-binomial distribution may often have little motivation other than statistical convenience. The double exponential family maximum likelihood method adds an extra parameter to the original one-parameter exponential family model that does not contain measurement error. By adding this parameter and allowing the overdispersion to be a function of covariates it is desired to adequately model the overdispersion associated with measurement error. This extension to the one-parameter model provides a natural motivation for the proposed method.

Both of these approximations are feedback models that use information about β contained in the variance of y_i given z_i to estimate β. Also both of these approximations use the correct mean and variance but assume different higher order moments from the true distribution of y_i given z_i and from each other. It is desired to know what effect each of these situations has on the estimation of β.

Iteratively weighted least squares (IWLS) is neither a feedback model nor does it assume any higher order moments. Thus, it gives us a benchmark for determining if situations described in the above paragraph have an effect on estimation of β. Also, because IWLS is straightforward to use and understand, it will be considered a good choice for estimation if it is fairly efficient.

It is expected that the exact MLE would be the most efficient when the underlying model is correct. The use of the information about β contained in the variance of y_i given z_i may give Prentice's and the DEF methods the edge over IWLS in terms of efficiency.

The simulations results support these expectations. For small measurement error there is relatively small differences between the mean square errors (MSEs) and the standard deviations of the estimates. For larger measurement error the distinctions are greater. The Lognormal MLE method has the smallest MSE and standard deviation of the estimate and IWLS has the largest. The differences between the four estimators which account for measurement error were quite small relative to the differences of each with the naive method.

Prentice's method estimates the Lognormal MLE result very closely for the identity link. The bias, MSE, standard deviation of the estimate and average standard error for these two methods are very close even for the largest measurement error standard deviation. The DEF method does less well than Prentice's method in approximating (8.1) for the identity link. For the simulation using the logit link, the DEF method does slightly better at this than Prentice's method.

Another important criterion for estimators, in addition to efficiency, is robustness. This is particularly important here since it is very difficult to check distributional assumptions with this kind of data. It is not expected that the exact MLE method is very robust. It is likely that the feedback models may be susceptible to outliers. Because iteratively weighted least squares is not based on a particular distribution and does not use information about β contained in the variability of y_i given z_i, it can be expected to be the most robust of this group.

A limited study of robustness was carried out with simulations here. The simulated situations from section 7.3 use estimators based on incorrect assumptions about the data. In these particular situations we see that the bias is very similar for the DEF method, IWLS and Prentice's method and even for the Lognormal MLE method. This could be due to the confounding discussed in section 7.3. The distribution for y_i given z_i is incorrect for the Lognormal MLE, and $E(x_i|z_i)$ and $Var(x_i|z_i)$ are incorrect for the remaining three methods. Thus, the situations considered have not allowed us to distinguish much between methods in terms of robustness. More work is needed in this area to determine how robust the methods are.

8.3 Other Issues

1) Accuracy of the Standard Errors. From the results of the simulation and from section 5.5, it seems that the DEF method is estimating the true variability fairly well in the situations considered. The tendency for all of the methods in the simulation study is that the average (asymptotic) standard error underestimates the standard deviation of the estimate.

2) Requirement that $E(x_i|z_i)$ and $Var(x_i|z_i)$ are known. In all that is done with the DEF method (and for IWLS and Prentice's method, as well), it has been assumed that $E(x_i|z_i)$ and $Var(x_i|z_i)$ are

known. There are few situations where it is convenient to estimate these values as presented in section 5.1. In general, they are not easy to obtain in practice. Considerable computation is necessary to arrive at the estimates in Table 1 for the chromosome aberration data set. In order to make all of these methods more practical, it would be desirable to know more situations where these values could be estimated easily and to know the consequences of using these estimates. In the simulation of section 7.2, the use of estimates of the nuisance parameters seems to do very well with only slightly higher variation for large measurement error.

BIBLIOGRAPHY

Armstrong, B., (1985), "Measurement Error in the Generalized Linear Model," *Communications in Statistics-Simulation and Computation*, 14, 529-544.

Baker, R. J., and Nelder, J. A., (1978), *The GLIM System: Release 3*, Oxford: Numerical Algorithms Group.

Berkson, J., (1950), "Are There Two Regressions?," *Journal of the American Statistical Association*, 45, 164-180.

Breslow, N. E., (1984), "Extra-Poisson Variation in Log-linear Models," *Applied Statistician*, 33, 38-44.

Carroll, R. J., and Rupert *D*., (1982), "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics*, 10, 429-441.

Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. T., and Abbott, R. D., (1984), "On Errors-in-Variables for Binary Regression Models," *Biometika*, 71, 19-25.

Cox, D. R., (1983), "Some Remarks on Overdispersion," *Biometrika*, 70, 269-274.

Crowder, M. J., (1978), "Beta-binomial Anova for Proportions," *Applied Statistician*, 27, 34-37.

Efron, B., (1986), "Double Exponential Families and Their Use in Generalized Linear Regression," *Journal of the American Statistical Association*, 81, 709-721.

Fuller, W. A., (1987), *Measurement Error Models*, New York: John Wiley & Sons.

Gilbert, E. S. (1985), "Some Effects of Random Dose Measurement Errors on Analysis of Atomic Bomb Survivor Data," *Radiation Research*, 98, 591-606 (RERF TR 12-82).

Greenwood, M., (1949), "The Infectiousness of Measles," *Biometrika*, 36, 1-8.

Hinde, J., (1982), "Compound Poisson regression models," In GLIM82: *Proceedings of the International Conference on Generalized Linear Models*. (R. Gilchrist, ed.), pp. 109-121. New York: Springer. (Lecture Notes on Statistics, Vol. 14)

Irwin, O. J., (1954), "A Distribution Arising in the Study of Infectious Diseases," *Biometrika*, 41, 266-268.

Iso, H., (1989), "Serum Cholesterol Levels and Six-Year Mortality From Stroke in 350,977 Men Screened For the Multiple Risk Factor Intervention Trial," *The New England Journal of Medicine*, 320, 904-910.

Johnson, N. L., and Kotz, S., (1987), *Continuous Univariate Distributions-1*, New York: John Wiley & Sons.

Kendall, M. G., and Stuart, A., (1961), *The Advanced Theory of Statistics, Volume 2*, New York: Hafner.

McCullagh, P., (1983), "Quasi-Likelihood Functions," *The Annals of Statistics*, 11, 59-67.

McCullagh, P., and Nelder J. A., (1983), *Generalized Linear Models*, London: Chapman & Hall.

Nelder, J. A., and Pregibon, D., (1987), "An Extended Quasi-Likelihood Function," *Biometrika*, 74, 221-232.

Otake, M., and Prentice, R. L., (1984), "The Analysis of Chromosomally Aberrant Cels Based on Beta-Binomial Distribution," *Radiation Research*, 98, 456-470.

Pierce, P. A., Stram, D. M., and Vaeth, M., (1989), "Allowing For Random Errors in Radiation Exposure Estimates for the Atomic Bomb Survivor Data," *RERF Technical Report* 2-89.

Pregibon, D., (1984), Review of *Generalized Linear Models*, by P. McCullagh and J. Nelder, *The Annals of Statistics*, 12, 1589-1596.

Prentice, R. L., (1986), "Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors," *Journal of the American Statistical Association*, 81, 321-327.

Schafer, D. W., (1987), "Covariate Measurement Error in Generalized Linear Models," *Biometrika*, 74, 385-91.

Schafer, D. W., (1988), "Multiplicative Measurement Errors in Explanatory Variables," *Oregon State University Technical Report*, 123.

Stefanski, L. A., (1985), "The Effects of Measurement Error on Parameter Estimation," *Biometrika*, 72, 583-92.

Stefanski, L. A., and Carroll, R. J., (1985), "Covariate Measurement Error in Logistic Regression," *The Annals of Statistics*, 13, 1335-1351.

Wedderburn, R. W. M., (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

White, H., (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

Whittemore, A. S., and Keller, J. B., (1988), "Approximations for Regression With Covariate Measurement Error," *Journal of the American Statistical Association*, 83, 1057-1066.

Williams, D. A., (1982), "Extra-binomial Variation in Logistic Linear Models," *Applied Statistician*, 31, 144-148.

APPENDICES

APPENDIX A

Data Sets

Chromosome Aberration Data: Number of Cells With Aberrations Out of 100 Cells Examined Per Individual.

Number of Aberrant Cells	Measured Radiation, z, (in rads*100)						
	0	0-1	1-2	2-3	3-4	4-5	5+
0	139	20	23	2	1		3
1	66	23	12	2	1		1
2	35	6	20	5	1	1	2
3	17	7	23	5	1	1	2
4	3	3	6	3	3	3	2
5	2	2	12	14	3	4	
6	1	5	12	3	3	1	1
7		2	12	2	3		
8			5	4	3	2	1
9		1	2	3	3	1	
10			5	5	3	3	
11			1	2	1	1	1
12				2	1	1	1
13				7	3	2	1
14		1	1	3	2	2	1
15			1	1		1	
16			2	2		1	3
17			1	2	2	1	4
18					4		
19					2		1
20					1	1	1
21				1			1
22					1	1	1
23							1
24				1			1
25				1		1	1
26-27							
28						1	
29						1	1
30-33							
34							1
35-36							
37						1	
38-39							
40							1
41							
42					1		1

Heart Disease Data: Deaths During Six Year Follow-Up and Age-Adjusted Rate of Death from Coronary Heart Disease, All Strokes, All Cardiovascular Disease, and All Causes, According to Serum Cholesterol Levels, in 350-977 Men, 35 to 57 Years of Age.

Serum Cholesterol mg/dl	Number at Risk	Actual Deaths	Age-adjusted Deaths per 10,000
< 140	5,062	9	25.5
140-159	16,123	33	24.6
160-179	42,276	122	32.9
180-199	65,381	227	36.7
200-219	76,570	398	52.1
220-239	61,856	433	67.5
240-259	41,332	338	77.8
260-279	23,749	291	116.1
280-299	10,445	138	124.4
> 300	8,183	135	160.3

APPENDIX B

Computer Code for
Binomial Simulation Using the Logit Link

```
program SIMULATION_of_the_Logit_Link(input,output);

const  n=100; m=100;              {<-------- n=#of obs, m=binomial sample size}
      mean_logx=4.0;  var_logx=0.5;  mean_logd=0.0;
      true_b1=0.05;  true_b2=0.0075;           {<-- True intercept and slope}
      maxparameters=2;

type vector = array[1..n] of real;
    dim = array[1..maxparameters] of real;
    matrix=array[1..maxparameters] of array[1..maxparameters] of real;
    mat1 = array[1..maxparameters] of array[1..1] of real;
    methods = (naive,adj_naive,naive_zs,IWLS,Exact_MLE,prentice,def1);
    vector2 = array[1..n] of integer;

var true_p,p,th,d,g,z,zs,vs,y : vector; y2 : vector2;
   r,j,pn,iterations,n1,q1,q2,q : integer;
   b,seb,t,individ1,dthdb,dpdb,dden,dddb : dim;
   score,product : mat1;
   info,individ2,covar : matrix;
   initial_b1,initial_b2,b2,c1,c2,var_est,var_logd,R : real;
   dldth,dldp,d2ldp2,zbar,dldg,d2ldg2,d2ldgdp : real;
   method : methods;  rel_eff_of_IWLS_to_DEF1 : real;
   beta2,sum_b2,sum_b22,sse,se_b2,sum_se,sum_se2 : array[methods] of real;
   ave_estimate,bias,sd_est,ave_se,sd_se,mse : array[methods] of real;
   total,time1 : array[methods] of integer;
   start,stop,start2,stop2,time2,hours,min,sec : integer;
   measurement_error : real;
   check : boolean;
   sumf,sumb1,sumb2,sumb12,sumb22,sumb1b2 : real;
   max_x : integer;

function power(base,exponent : real) : real;   {<-raises base to power exponent}
begin
 if (base=0.0) then power:=0.0
  else power:=exp(exponent*ln(base));
end;

procedure initialize_sums;
begin
 for method:=naive to def1 do begin
  sum_b2[method]:=0.0; sum_b22[method]:=0.0; sse[method]:=0.0;
  se_b2[method]:=0.0; sum_se[method]:=0.0; sum_se2[method]:=0.0;
  total[method]:=0;  time1[method]:=0;
 end;
 var_logd:=measurement_error*measurement_error;
```

```
 R:=var_logx/(var_logx+var_logd);                         {<-- Reliability ratio}
end;

{ rnorm generates n obs. from a Normal(mean,variance) }

procedure rnorm(mean,variance : real; var  nm : vector);
var  a,b,k,rnorm1,rnorm2 : real; i : integer;
begin
 i:=1;
   while (i <= n) do begin
    a:=2.0*random(q1)-1.0;  b:=2.0*random(q2)-1.0;
    k:=a*a+b*b;
    if (k <= 1.0) and (k<>0.0) then begin
      rnorm1:=(a*sqrt(-2.0*ln(k)/k))*sqrt(variance) +mean;
      rnorm2:=(b*sqrt(-2.0*ln(k)/k))*sqrt(variance) +mean;
      nm[i]:=rnorm1;  nm[i+1]:=rnorm2;
      i:=i+2;
    end; {*if*}
   end; {*while*}
 end; {*proc*}

{ rbinomial generates obs from binomial(m,true_p[i]) }

procedure  rbinomial; var  i,j,count : integer;
begin
 for i:=1 to n do begin
  count:=0;
  for j:=1 to m do if (random(q1) < true_p[i])
   then count:=count+1;
  y[i]:=count/m; y2[i]:=count;
 end; {*i*}
end; {*proc*}

{ Generates data and calculates E(x|z) and Var(x|z) }

procedure generate_data;
var lx,x,lz,lh : vector;
    i : integer;  mu,ada,sigma : real;
begin
 rnorm(mean_logx,var_logx,lx);  rnorm(mean_logd,var_logd,lh);
 for i:= 1 to n do begin
  x[i]:=exp(lx[i]);
  true_p[i]:=exp(true_b1+true_b2*x[i])/(1+exp(true_b1+true_b2*x[i]));
  lz[i]:=lx[i]+lh[i]; z[i]:=exp(lz[i]);   {* Measurement Error in the z's *}
 end;{i}
 mu:=mean_logx;  sigma:=var_logx;          {<-- setting the values for the }
 ada:=mu-R*mu; zbar:=mean_logx;            {     known nuisance parameters }
 for i:= 1 to n do begin
  zs[i]:=power(z[i],R)*exp(ada+0.5*(1-R)*sigma);                   {*E(x|z)*}
  vs[i]:=power(z[i],2*R)*exp(2*ada+(1-R)*sigma)*(exp((1-R)*sigma)-1);{*V(x|z)*}
 end; {i}
 rbinomial;
```

```
end; {proc}

procedure set_derivatives_0; var j,k : integer;
begin
 for j:=1 to pn do score[j,1]:=0.0;
 for j:=1 to pn do  for k:=1 to pn do info[j,k]:=0.0;
end;

procedure update_p_theta; var i : integer; z1 : real;
begin
 b2:=b[2]*b[2];                              {<-- slope squared   }
 for i:= 1 to n do begin
  if (method=naive) then z1:=z[i]  else z1:=zs[i];

                              { the inverse of the logit link for p }
  p[i]:=exp(b[1]+b[2]*z1)/(1+exp(b[1]+b[2]*z1)); {<----' }

  if (method=def1) then
   th[i]:=1+((m-1)*b2*vs[i])*p[i]*(1-p[i]);      {<-- reciprocal of dispersion }

  if (method=prentice) then begin
   d[i]:=b2*vs[i]*(p[i]*(1-p[i]));
   g[i]:=d[i]/(1-d[i]);
  end; {if prentice}
 end; {i}
 for j:=1 to pn do t[j]:=b[j];               {<-- stores last value of beta}
end; {proc}

procedure loglikder(i : integer);     {<--  derivatives for the beta-binomial}
var k : integer;
   s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12,k1,k2,k3 : real;
begin
 s1:=0.0; s2:=0.0; s3:=0.0; s4:=0.0; s5:=0.0; s6:=0.0; s7:=0.0;
 s8:=0.0; s9:=0.0; s10:=0.0; s11:=0.0; s12:=0.0;
    if y2[i]>0 then begin
     for k:= 0 to y2[i]-1 do begin
      k1:= 1/(p[i]+k*g[i]);
      s1:= s1 + k1;
      s3:= s3 - k1*k1;
      s5:= s5 + k*k1;
      s8:= s8 - k*k*k1*k1;
      s11:= s11 - k*k1*k1;
     end; {k loop}
    end; {check for y=0}

    if y2[i]<100 then begin
     for k:= 0 to m-y2[i]-1 do begin
      k2:= 1/((1-p[i])+g[i]*k);
      s2:= s2 - k2;
      s4:= s4 - k2*k2;
      s6:= s6 + k*k2;
      s9:= s9 - k*k*k2*k2;
```

```
      s12:= s12 + k*k2*k2;
     end; {k loop}
    end; {check for y=100}

    for k:= 0 to m-1 do begin
     k3:= k/(1+k*g[i]);
     s7:= s7 - k3;
     s10:= s10 + k3*k3;
    end;

   dldp:=s1+s2;  dldg:=s5+s6+s7;           {    The derivatives of the       }
   d2ldp2:= s3 + s4;                       {<-- log-likelihood function with}
   d2ldg2:= s8 + s9 + s10;                  {    respect to p and gamma      }
   d2ldgdp:= s11 + s12;
end; {*proc*}

procedure firstderivative(i : integer);
var w,ths,dthsdth,dgdd : real;
begin
 if (method=def1) then begin              {*<---------def1 1st derivatives*}

  if (y[i]<>0.0) then w:=y[i]  else w:=y[i] + 0.0001;   {<-- checks for zeros}
  if (y[i]=1.0) then w:=y[i] - 0.0001;                 {<-- checks for ones }
  ths:=1/th[i];  dthsdth:=-1/(th[i]*th[i]);
   dldth:=1/(2*ths)+m*w*ln(p[i])+(m-m*w)*ln(1-p[i])-m*w*ln(w)-(m-m*w)*ln((1-w));
  dldp:=m*ths*(y[i]-p[i])/(p[i]*(1-p[i]));

  c1:=p[i]*(1-p[i]); c2:=c1*(1-exp(b[1]+b[2]*zs[i]))/(1+exp(b[1]+b[2]*zs[i]));

  dpdb[1]:=c1;                           {<-- derivative of p w/r beta0}
  dpdb[2]:=zs[i]*c1;                      {<-- derivative of p w/r beta1}

  dthdb[1]:=(m-1)*b2*vs[i]*c2;                  {<-- der. of theta  w/r beta0}
  dthdb[2]:=(m-1)*vs[i]*(2*b[2]*c1+b2*zs[i]*c2);  {<-- der. of theta  w/r beta1}

{1st derivative contributions for an individual}

  for j:=1 to pn do
   individ1[j]:=dldp*dpdb[j] + dldth*dthsdth*dthdb[j];
 end; {*def1*}

 if (method=prentice) then begin        {*<---------prentice 1st derivatives*}
  dgdd:=1/((1-d[i])*(1-d[i]));
  loglikder(i);
    c1:=p[i]*(1-p[i]); c2:=c1*(1-exp(b[1]+b[2]*zs[i]))/(1+exp(b[1]+b[2]*zs[i]));
  dpdb[1]:=c1;
  dpdb[2]:=zs[i]*c1;
  dddb[1]:=b2*vs[i]*c2;                    {<-- der. of delta  w/r beta0}
  dddb[2]:=vs[i]*(2*b[2]*c1+b2*zs[i]*c2);       {<-- der. of delta  w/r beta1}

{1st derivative contributions for an individual}
```

```
  for j:=1 to pn do  individ1[j]:=dldp*dpdb[j] + dldg*dgdd*dddb[j];
 end; {*prentice*}
end; {proc}

procedure secondderivative(i : integer);
var z,v,d2ldth2,d2ldpdth,ths,dthsdth,dgdd : real;
   d2thsdth2,d2gdd2,k1,k2,c3 : real;
   d2thdb,d2pdb,d2ddb : matrix; j,k : integer;
begin
 if (method=def1) then begin                    {<------- def1 2nd derivatives}
  z:=zs[i];  v:=vs[i];
  k1:=exp(b[1]+b[2]*z);  k2:=exp(b[1]+b[2]*z)*exp(b[1]+b[2]*z);
  c3:=c1*(1-4*k1+k2)/((1+k1)*(1+k1));

  ths:=1/th[i];  dthsdth:=-1/(th[i]*th[i]); d2thsdth2:=2/(th[i]*th[i]*th[i]);

  d2pdb[1,1]:=c2;
  d2pdb[2,2]:=z*z*c2;                       {<-- derivatives of p w/r beta}
  d2pdb[1,2]:=z*c2;
  d2pdb[2,1]:=z*c2;

  d2thdb[1,1]:=(m-1)*v*b2*c3;               {<-- derivatives of theta w/r beta}
  d2thdb[2,2]:=(m-1)*v*(2*c1+4*b[2]*z*c2+b2*z*z*c3);
  d2thdb[1,2]:=(m-1)*v*(2*b[2]*c2+b2*z*c3);
  d2thdb[2,1]:=d2thdb[1,2];

  d2ldth2:=-1/(2*ths*ths);
  d2ldp2:=-m*y[i]*ths/(p[i]*p[i]) - (m-m*y[i])*ths/((1-p[i])*(1-p[i]));
  d2ldpdth:=m*y[i]/p[i]-(m-m*y[i])/(1-p[i]);
  for j:=1 to pn do  for k:=1 to pn do

{2nd derivative contributions for an individual}

  individ2[j,k]:=(d2ldp2*dpdb[k]+d2ldpdth*dthsdth*dthdb[k])*dpdb[j]
             +dldp*d2pdb[j,k]
        +(d2ldpdth*dpdb[k]+d2ldth2*dthsdth*dthdb[k])*dthsdth*dthdb[j]
           +dldth*(dthsdth*d2thdb[j,k] + d2thsdth2*dthdb[j]*dthdb[k]);
end; {*def1*}

if (method=prentice) then begin          {*<---------prentice 2nd derivatives*}
 dgdd:=1/((1-d[i])*(1-d[i]));  d2gdd2:=2/((1-d[i])*(1-d[i])*(1-d[i]));
 z:=zs[i];  v:=vs[i];
 k1:=exp(b[1]+b[2]*z);  k2:=exp(b[1]+b[2]*z)*exp(b[1]+b[2]*z);
 c3:=c1*(1-4*k1+k2)/((1+k1)*(1+k1));

 d2pdb[1,1]:=c2;
 d2pdb[2,2]:=z*z*c2;                        {<-- derivatives of p w/r beta}
 d2pdb[1,2]:=z*c2;
 d2pdb[2,1]:=z*c2;

 d2ddb[1,1]:=v*b2*c3;                       {<-- derivatives of delta w/r beta}
 d2ddb[2,2]:=v*(2*c1+4*b[2]*z*c2+b2*z*z*c3);
```

```
  d2ddb[1,2]:=v*(2*b[2]*c2+b2*z*c3);
  d2ddb[2,1]:=d2ddb[1,2];

{2nd derivative contributions for an individual}

  for j:=1 to pn do  for k:=1 to pn do
   individ2[j,k]:=(d2ldp2*dpdb[k]+d2ldgdp*dgdd*dddb[k])*dpdb[j]
                 +dldp*d2pdb[j,k]
                  +(d2ldgdp*dpdb[k]+d2ldg2*dgdd*dddb[k])*dgdd*dddb[j]
                    +dldg*(d2gdd2*dddb[j]*dddb[k] + dgdd*d2ddb[j,k]);
 end; {*prentice*}
end; {proc}

{Finds the inverse of the information matrix}

procedure inverse; var temp : real; i,j,k : integer;
begin
 for i:=1 to pn  do begin for j:=1 to pn do begin
   if i=j then  covar[i,j]:=1.0
    else  covar[i,j]:=0.0; end; end;
 for i:=1 to pn do begin for j:=1 to pn do begin
   if j=1 then temp:=info[i,i];
   if temp=0.0 then writeln('**  warning singular matrix  **');
   info[i,j]:=info[i,j]/temp;  covar[i,j]:=covar[i,j]/temp; end;
  for k:=1 to pn do begin
   if (k<>i) then begin
    for j:=i to pn do  begin
     if i=j then temp:=info[k,i];
     info[k,j]:=-temp*info[i,j]+info[k,j]; end;
   for j:=1 to pn do  covar[k,j]:=-temp*covar[i,j]+covar[k,j];
   end; end; end; end;

{This does the matrix multiplication for the Newton-Raphson method}

procedure matrixmult(r1,c1,c2 : integer); var temp : real;
i,j,k : integer;
begin
 for i:=1 to r1 do
  for k:=1 to c1 do
   product[i,k]:=0.0;
 for i:=1 to r1 do begin
  for k:=1 to c1 do begin
   for j:=1 to c2 do begin
    temp:=covar[i,j]*score[j,k];
    product[i,k]:=product[i,k]+temp;
end; end; end; end; {proc}

{Using the Newton-Raphson method of updating estimates}

procedure estimates;
var i : integer;
begin
```

```
 inverse;
 matrixmult(pn,1,pn);
 for i:=1 to pn do b[i]:=t[i]-product[i,1];
end; {estimates}

procedure set_std_errors; var i : integer;
begin
 for i:=1 to pn do seb[i]:=sqrt(-covar[i,i]);
end;

function convergence(var old : dim; new : dim) : boolean;
var k : integer;
begin
 convergence:=FALSE;
 for k:=1 to pn do
  if (abs((old[k]-new[k])/new[k])>0.0001) then convergence:=TRUE;
end;

function b2_convergence : boolean;     {<-- checks for unusual answers like NaN}
begin
 b2_convergence:=FALSE;
 if (beta2[method]>-100.0) and (beta2[method]<100.0)
   and (se_b2[method]>0.0) and (se_b2[method]<100.0)
    then b2_convergence:=TRUE;
end;

{ Cummulates the individual contributions for the score vector and the }
{   information matrix                                                  }

procedure score_and_info; var j,k : integer;
begin
 for j:=1 to pn do score[j,1]:=score[j,1]+individ1[j];
 for j:=1 to pn do for k:=1 to pn do info[j,k]:=info[j,k]+individ2[j,k];
end;

{Sets initial guess for last 4 methods to the Naive_zs estimate}

procedure set_initial_guess; var i : integer;
begin
 b[1]:=initial_b1; b[2]:=initial_b2;
 for i:=1 to pn do t[i]:=0.0; iterations:=0;
end;

procedure weighted_least_squares;
var sumtop,sumbottom,sumwy,sumwzs,sumw,sumwzsy,sszs,sigma,bzsy,ssy : real;
   w,z1,p1,y1,ada : real;
   i : integer;
begin
 sumwy:=0.0; sumwzs:=0.0; sumw:=0.0; sumwzsy:=0.0; sszs:=0.0; ssy:=0.0;
 for i:=1 to n do begin
  if (method=naive) then z1:=z[i] else z1:=zs[i];
  if (method=naive) or (method=naive_zs) then
```

```
  if (iterations=0) then p1:=y[i] else p1:=p[i];  {<-- sets initial guess }
 if (method=IWLS) then p1:=p[i];
  if (p1=1.0) then p1:=0.9999;
  if (p1=0.0) then p1:=0.0001;
 ada:=ln(p1/(1-p1));
 y1:=ada + (y[i]-p1)/(p1*(1-p1));          {<--  working dependent variable}
 if (method=naive) or (method=naive_zs) then
  w:=m*p1*(1-p1) else                      {<--  usual binomial weight}
  w:=m*p1*(1-p1)/(1+(m-1)*b2*p1*(1-p1)*vs[i]); {<--  Armstrong's weight}
 sumwzsy:=sumwzsy+w*z1*y1; sumwy:=sumwy+w*y1;
 sumwzs:=sumwzs+w*z1; sumw:=sumw+w; sszs:=sszs+z1*w*z1;
 ssy:=ssy+y1*w*y1;
end; {*i*}

{y1 regressed on z1 with weights w}

sumtop:=sumw*sumwzsy-sumwzs*sumwy;  sumbottom:=sumw*sszs-sumwzs*sumwzs;
b[2]:=sumtop/sumbottom;
b[1]:=(sumwy*sszs-sumwzs*sumwzsy)/(sumw*sszs-sumwzs*sumwzs);
bzsy:=b[1]*sumwy+b[2]*sumwzsy;
sigma:=sqrt((ssy-bzsy)/(n-pn));
seb[1]:=1/sqrt(sumbottom/sszs);
seb[2]:=1/sqrt(sumbottom/sumw);
end; {*proc*}

{Iterative cycle for IWLS}

procedure IWLS_loop;
begin
 set_derivatives_0; set_initial_guess;
 while convergence(t,b) do begin
  update_p_theta;
  weighted_least_squares;
  iterations:=iterations+1;
 end; {*while*}
 update_p_theta;
 beta2[method]:=b[2]; se_b2[method]:=seb[2];
end; {proc}

{Iterative cycle for DEF, Prentice and EXACT MLE}

procedure estimation_loop;
var i : integer;
begin
 set_derivatives_0; set_initial_guess;
 while convergence(t,b) do begin
  update_p_theta;
  for i:=1 to n do begin
   firstderivative(i); secondderivative(i); score_and_info;
  end; {*i*}
  estimates;  set_derivatives_0; iterations:=iterations+1;
  if iterations>30 then for i:=1 to pn do t[i]:=b[i];
```

```
 end; {*while*}
 update_p_theta;  set_std_errors;
 beta2[method]:=b[2]; se_b2[method]:=seb[2];
end; {proc}

procedure naive_method;                    {<-- Regresses y on z using }
begin                                     {    binomal regession (ML) }
 method:=naive; pn:=2;
 start:=wallclock;
 IWLS_loop;
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
end;

procedure naive_zs_method;                 {<-- Regresses y on E(x|z) using }
begin                                     {    binomal regession (ML) }
 method:=naive_zs; pn:=2;
 start:=wallclock;
 IWLS_loop;
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
 initial_b2:=b[2];
 initial_b1:=b[1];
end;

procedure adj_naive_method; var e2,L0,L1,mean_x,mean_h : real;
begin
 method:=adj_naive;
 e2:=measurement_error*measurement_error;
 L1:=(exp(var_logx)-1)/(exp(e2/2)*(exp(var_logx+e2)-1));
 mean_x:=exp(mean_logx+var_logx/2);
 mean_h:=exp(e2/2);
 L0:=mean_x-L1*mean_x*mean_h;
 b[2]:=b[2]*(1/L1);             {<--Corrects for the bias in the Naive method}
 b[1]:=b[1]-b[2]*L0;
 beta2[method]:=b[2]; se_b2[method]:=(1/L1)*seb[2];
end;

procedure armstrong_method_IWLS;          {Iteratively weighted least squares}
begin
 method:=IWLS; pn:=2;
 start:=wallclock;
 IWLS_loop;
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
end; {proc}

procedure DEF_Method_1;     {*basic DEF method*} var i : integer;
begin
 start:=wallclock;
 method:=def1; pn:=2;
 estimation_loop;
 if (not b2_convergence) or (iterations>30) then begin
  write(method,'convergence problem ',r); check:=FALSE;
  writeln(iterations:5);
```

```
 end; {if}
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
end; {DEF}

procedure Prentice_Method;
begin
 start:=wallclock;
 method:=prentice; pn:=2;
 estimation_loop;
  if (not b2_convergence) or (iterations>30) then begin
   write(method,'convergence problem ',r); check:=FALSE;
   writeln(iterations:5);
  end; {if}
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
end; {Prentice_Method}

{The next 3 procedures are for the Lognormal MLE method}

{This procedure calculates the derivatives for the Lognormal MLE method}

procedure summations(f,fb1,fb2,f2b1,f2b2,f2b1b2,c : real;
                    j,step_size : integer;  check2 : boolean);
begin
  if (j=0) or (j=max_x) then begin
   sumf:=sumf+c*f;
   sumb1:=sumb1+c*fb1;
   sumb2:=sumb2+c*fb2;
   sumb12:=sumb12+c*f2b1;
   sumb22:=sumb22+c*f2b2;
   sumb1b2:=sumb1b2+c*f2b1b2;
   check2:=FALSE;
  end; {zero,max}
  if (j mod (2*step_size) = 0) and check2 then begin
   sumf:=sumf+c*f*2;
   sumb1:=sumb1+c*fb1*2;
   sumb2:=sumb2+c*fb2*2;
   sumb12:=sumb12+c*f2b1*2;
   sumb22:=sumb22+c*f2b2*2;
   sumb1b2:=sumb1b2+c*f2b1b2*2;
  end; {even}
  if (j mod (2*step_size) = (step_size)) then  begin
   sumf:=sumf+c*f*4;
   sumb1:=sumb1+c*fb1*4;
   sumb2:=sumb2+c*fb2*4;
   sumb12:=sumb12+c*f2b1*4;
   sumb22:=sumb22+c*f2b2*4;
   sumb1b2:=sumb1b2+c*f2b1b2*4;
  end; {odd}
end; {proc}

{Numerical integration for f(y|z) and derivatives for Lognormal MLE}
```

```
procedure numerical_integration(i : integer);
var mu,my,p,p_my,q_m_my,p_my_1,p_my_2,q_m_my_1,q_m_my_2 : real;
   x,k1,k2,k3,f,fb1,f2b1,fb2,f2b2,f2b1b2,c,e1,e2,e3,e4,e5 : real;
   j,step_size,old_step,last_change : integer; check2,check3 : boolean;
   dpdb1,dpdb2,d2pdb12,d2pdb22,d2pdb1b2 : real;
begin
 check3:=FALSE;
 sumf:=0.0; sumb1:=0.0; sumb2:=0.0; sumb12:=0.0; sumb22:=0.0; sumb1b2:=0.0;
 mu:=mean_logx*(1-R)+R*ln(z[i]);
 my:=m*y[i];
 j:=0; step_size:=2; old_step:=2; last_change:=2000;
 while j<=max_x do begin
  check2:=TRUE;
  if (j=0) then x:=0.0001 else x:=j;
  p:=exp(b[1]+b[2]*x)/(1.0+exp(b[1]+b[2]*x));
  p_my:=power(p,my);   p_my_1:=power(p,my-1);   p_my_2:=power(p,my-2);
  q_m_my:=power(1.0-p,m-my); q_m_my_1:=power(1.0-p,m-my-1);
  q_m_my_2:=power(1.0-p,m-my-2);
  k1:=sqrt(2*arctan(1)*4*R)*measurement_error;
  k2:=-0.5/(R*measurement_error*measurement_error);
  k3:=(1/(x*k1)*exp(k2*(ln(x)-mu)*(ln(x)-mu)));
  f:=p_my*q_m_my*k3;
   e1:=my*p_my_1*q_m_my;   e2:=(m-my)*p_my*q_m_my_1;

  c1:=p*(1.0-p); c2:=c1*(1.0-exp(b[1]+b[2]*x))/(1+exp(b[1]+b[2]*x));

  dpdb1:=c1;                           {<-- derivative of p w/r beta0}
  dpdb2:=x*c1;                          {<-- derivative of p w/r beta1}

  fb1:=k3*(e1-e2)*dpdb1;
  fb2:=k3*(e1-e2)*dpdb2;
   e3:=my*(my-1)*p_my_2*q_m_my;   e4:=my*(m-my)*p_my_1*q_m_my_1;
   e5:=(m-my)*(m-my-1)*p_my*q_m_my_2;

  d2pdb12:=c2;
  d2pdb22:=x*x*c2;                {<-- derivatives of p w/r beta}
  d2pdb1b2:=x*c2;

  f2b1:=k3*((e3-2*e4+e5)*dpdb1*dpdb1+(e1-e2)*d2pdb12);
  f2b2:=k3*((e3-2*e4+e5)*dpdb2*dpdb2+(e1-e2)*d2pdb22);
  f2b1b2:=k3*((e3-2*e4+e5)*dpdb1*dpdb2+(e1-e2)*d2pdb1b2);
  if (f<=0.0000000000000000000000001) and (j mod 40 = 0) then
   if (j>199) then step_size:=20;             {larger step size for integ.}
  if (step_size=20) and (j mod 100 = 0) then
   if (j>=last_change+100) then step_size:=100;    {larger step size for integ.}
  if (step_size=old_step) then c:=step_size/3   {width of block being integrat.}
   else begin
    c:=(step_size+old_step)/6;
    last_change:=j;
   end;
  summations(f,fb1,fb2,f2b1,f2b2,f2b1b2,c,j,step_size,check2);
  j:=j+step_size;                          {set next pt of integrat.}
```

```
  old_step:=step_size;
 end; {while}
end; {proc}

{Computes score vector and information for Lognormal MLE method}

procedure exact_derivatives; var sum2 : real; i : integer;
begin
 max_x:=600;
 if (max_x mod 2 = 1) then max_x:=max_x-1;
 for i:=1 to n do begin
  numerical_integration(i);
  score[1,1]:=score[1,1]+(sumb1/sumf);
  score[2,1]:=score[2,1]+(sumb2/sumf);
   sum2:=sumf*sumf;
  info[1,1]:=info[1,1]+(sumb12*sumf-sumb1*sumb1)/sum2;
  info[2,2]:=info[2,2]+(sumb22*sumf-sumb2*sumb2)/sum2;
  info[1,2]:=info[1,2]+(sumb1b2*sumf-sumb1*sumb2)/sum2;
 end; {i}
 info[2,1]:=info[1,2];
end; {proc}

procedure EXACT_MLE; var j : integer;           {<-- Lognormal MLE method}
begin
 method:=exact_MLE; pn:=2;
 start:=wallclock;
 set_derivatives_0; set_initial_guess;
 while convergence(t,b) do begin
  for j:=1 to pn do t[j]:=b[j];
  exact_derivatives;
  estimates; set_derivatives_0; iterations:=iterations+1;
  if iterations>30 then for j:=1 to pn do t[j]:=b[j];
 end; {while}
 beta2[method]:=b[2]; se_b2[method]:=sqrt(-covar[2,2]);
 if (not b2_convergence) or (iterations>30) then begin
  writeln(method,'convergence problem ',r); check:=FALSE; end;
 stop:=wallclock; time1[method]:=time1[method]+(stop-start);
end; {proc}

{Cummulates the summary statistics}

procedure sum_statistics;
begin
 for method:= naive to def1 do begin
  if b2_convergence then begin        {*<------checks for convergence of b2*}
   sum_b2[method]:=sum_b2[method] + beta2[method];
   sum_b22[method]:=sum_b22[method] + beta2[method]*beta2[method];
   sse[method]:= sse[method] + (beta2[method]-0.0075)*(beta2[method]-0.0075);
   sum_se[method]:= sum_se[method] + se_b2[method];
   sum_se2[method]:= sum_se2[method] + se_b2[method]*se_b2[method];
   total[method]:= total[method] + 1;
  end; {*b2_convergence*}
```

```
 end; {*method*}
end; {*proc*}

{Calculates the summary statistics for the 1000 data sets}

procedure calc_summary_stats;
begin
 for method:= naive to def1 do begin
  n1:=total[method];
  if n1>0 then begin
   ave_estimate[method]:=sum_b2[method]/n1;
   ave_se[method]:= sum_se[method]/n1;
   mse[method]:=sse[method]/n1;
  end; {*if 0*}
  if n1>1 then begin
   var_est:=(sum_b22[method]-sum_b2[method]*sum_b2[method]/n1)/(n1-1);
   sd_est[method]:=sqrt(var_est);
   sd_se[method]:=sqrt((sum_se2[method]
                -sum_se[method]*sum_se[method]/n1)/(n1-1));
  end; {*if 1*}
 end; {*method*}
rel_eff_of_IWLS_to_DEF1:=sd_est[IWLS]*sd_est[IWLS]/(sd_est[def1]*sd_est[def1]);
end; {*proc*}

{Outputs results}

procedure print_table;
begin
writeln;
writeln; write('Measurement Error = ',measurement_error:4:2); writeln('  R= ',R:5:3); writeln;
writeln('          Average of  Mean Square    SD of      Average       SD of');
writeln(' Method   Estimate      Error       Estimate      SE            SE');
writeln('------------------------------------------------------------------------------');
for method:= naive to def1 do begin
write(method:9,ave_estimate[method]:10:7,mse[method]:16:12);
write(sd_est[method]:13:9,ave_se[method]:14:11,sd_se[method]:18:14);
writeln;
end; {*method*}
writeln;
writeln; write('Relative efficiency of IWLS to DEF1 =  ');
 writeln(rel_eff_of_IWLS_to_DEF1:9:6); writeln;
writeln;
for method:=naive to def1 do begin
hours:=time1[method] div 3600;
min:=(time1[method] mod 3600) div 60;
sec:=(time1[method] mod 3600) mod 60;
if method<>adj_naive then
writeln(method:9,' took ',hours:4,' hours ',min:4,' minutes ',sec:4,' seconds');
end;
end; {*proc*}

procedure print_fixed_constants;
```

```
begin
writeln('Data sets = 1000 and n=100 and m=100');
writeln('Mean of the true logx = ',mean_logx:5:3,'  Variance of true logx = ',var_logx:6:4);
writeln; writeln('Starting seed=',q2:15);
end; {*proc*}

begin
 q1:=(wallclock mod 1000000)*1217;                { Initializes the starting}
 q2:=abs((q1*8192-67099547*trunc(q1*8192/67099547)));{<- values for the random }
 q:=seed(q2);                                { number generator        }
 print_fixed_constants;
 measurement_error:=0.2;
while  measurement_error<=0.5 do begin
 start2:=wallclock;
 initialize_sums; r:=1;
 while (r < 1001) do begin
  check:=TRUE;
  generate_data;

  naive_method;
  adj_naive_method;
  naive_zs_method;
  armstrong_method_IWLS;
  DEF_Method_1;
  Prentice_Method;
  EXACT_MLE;

  if check then begin              {<-- Checks the convergence of all methods}
   r:=r+1;
   sum_statistics;
  end;{*if*}
 end; {*r*}

calc_summary_stats;
print_table;
stop2:=wallclock;
time2:=stop2-start2;
hours:=time2 div 3600;
min:=(time2 mod 3600) div 60;
sec:=(time2 mod 3600) mod 60;
writeln('total time was ',hours:4,' hours ',min:4,' minutes ',sec:4,' seconds');
measurement_error:=measurement_error+0.1;
end; {*measurement_error*}
writeln('n= ',n:3,'   m= ',m:3,'  Monte Carlo Runs=',r-1:3);
end.
```