AN ABSTRACT OF THE THESIS OF

Roger Higdon for the degree of Doctor of Philosophy in Statistics presented on September 23 1998. Title: Application and Computation of Likelihood Methods for Regression with Measurement Error.

Abstract approved: _____

Daniel W. Schafer

This thesis advocates the use of maximum likelihood analysis for generalized regression models with measurement error in a single explanatory variable. This will be done first by presenting a computational algorithm and the numerical details for carrying out this algorithm on a wide variety of models. The computational methods will be based on the EM algorithm in conjunction with the use of Gauss-Hermite quadrature to approximate integrals in the E-step. Second, this thesis will demonstrate the relative superiority of likelihood-ratio tests and confidence intervals over those based on asymptotic normality of estimates and standard errors, and that likelihood methods may be more robust in these situations than previously thought. The ability to carry out likelihood analysis under a wide range of distributional assumptions, along with the advantages of likelihood ratio inference and the encouraging robustness results make likelihood analysis a practical option worth considering in regression problems with explanatory variable measurement error.

APPLICATION AND COMPUTATION OF LIKELIHOOD METHODS FOR REGRESSION WITH MEASUREMENT ERROR

by

Roger Higdon

A THESIS

Submitted to

Oregon State University

in partial fulfillment of
the requirement for the
degree of

Doctor of Philosophy

Presented September 23, 1998
Commencement June 1999

Doctor of Philosophy thesis of Roger Higdon presented on September 23, 1998

APPROVED:

Major Professor, representing Statistics

Chair of Department of Statistics

Redacted for privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes the release of my thesis to any reader upon request.

Redacted for privacy

Roger Higdon, Author

ACKNOWLEDGMENTS

I would like to take the time to thank a few of the people who made this thesis possible. First, is my advisor Dan Schafer, without whose help, knowledge, and ideas this thesis would not have been written. Thanks big guy! Next, I would like to thank everybody at the Oregon State University, Department of Statistics for putting up with me for six years and teaching me all the statistics that made this Ph.D. possible. I really enjoyed my stay and I will miss everyone. Most importantly, I would like to thank Mom, who is most responsible for who I am and what I have accomplished. Finally, I would like to thank Erica, who has made the last year and a half of this degree a lot more enjoyable than it would have been otherwise.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX TABLES

APPLICATION AND COMPUTATION OF LIKELIHOOD METHODS FOR REGRESSION WITH MEASUREMENT ERROR

1. INTRODUCTION

1.1 Purpose

Although many authors have addressed the problem of measurement error in explanatory variables, a relatively small amount of attention has been given to likelihood methods in comparison to methods based on moment assumptions. There are several reason for this. First, likelihood methods generally involve much greater computational difficulties than moment based methods. Second, there is a great deal of uncertainty about the robustness of likelihood methods because of their strong distributional assumptions. And third, there is a general belief that methods based on weaker moment assumptions may perform just as well in practice, although there is very little work in the literature to support the last claim. The purpose of this thesis is to address some of these deficiencies in the literature. First, this thesis will provide a framework and computational methods to apply likelihood methods to a broad range of generalized regression models where a single explanatory variable is measured with error. Second, it will provide some insight into the efficiency and robustness of likelihood methods relative to moment based methods through a series of simulations. This thesis will show there are some definite advantages of likelihood analysis over simpler moment based methods, and that the computational obstacles to likelihood analysis can be overcome. Therefore, this thesis advocates that likelihood analysis should at least be considered in practice in regression problems involving explanatory variable measurement error.

1.2 Background

1.2.1 Measurement Error

In most regression models of a response variable, Y, on an explanatory variable, X, it is generally assumed that the explanatory variable is measured exactly, without error. However, there are many situations where the explanatory variable, X, is not known exactly; rather a measurement or surrogate of the explanatory variable, W, is taken in its place. If one uses the regression of Y on W to estimate the regression of Y on X, this generally results in biased estimates for the regression coefficients of the regression of Y on X. If this bias is large enough, inference based on these estimates will be unreliable. Therefore, it is necessary in these situations to use methods that account for this measurement error.

In order to adjust for the measurement error it is necessary to make some assumptions about the nature of it. This thesis will deal primarily with the "classical" measurement error problem, where the observed explanatory variable, W, can be thought of as a combination of the true explanatory variable, X, and measurement error, U. The simplest models will assume that this measurement error is additive (possibly on the log scale), where $W = X + U$ (or $W = XU$) with X and U taken to be independent, and U having mean zero (or median 1 in the multiplicative case). The usual way one might suspect this type of measurement error to occur would be from using an inaccurate device to measure a physical quantity. However, there are many other ways; for example, in a dietary study a survey of one days food consumption for a patient may be used to estimate saturated fat intake. Errors could occur here not only from errors in a patient's recollection about their consumption, but also from the fact that a single day consumption varies from a patient's overall diet. Another form of measurement error is the Berkson measurement error model (Berkson, 1950). In this case the observed explanatory variable is fixed by the experimenter, but the true explanatory variable is a combination of the observed explanatory variable and random error, $X = W + U$ with W and U independent in this

case. An example of this type of measurement error might occur in a greenhouse experiment where the temperature of the greenhouse might be set by the experimenter using a thermostat but the actual temperature within a greenhouse varies due to the inaccuracy of the heating device. The use of W in place of X as the explanatory variable generally does not result in additional bias for estimates for the coefficients of the regression of Y on X, but there is bias in the estimate of the regression variance when the measurement error is of this type.

1.2.2 Example

An example of explanatory variable measurement error in regression comes from a study by Clayton (1991), in which the ratio of polyunsaturated to saturated fat intake (P/S) was related to death by heart disease. This ratio was measured on 336 male subjects by a one week dietary survey. The survey was repeated for a subset of 76 subjects six months later. In this case Y is 1 if the subject died from heart disease during the study and 0 otherwise, and X may be thought of as the long term mean of the log of P/S (X is logged to better fit the structural model). In this study X cannot be observed rather a measurement, W, is used by calculating log P/S from the one week dietary survey. The measurement error, U, represents only the variation in diet from week to week, since the measurement of log P/S from the survey was done quite accurately. To use structural likelihood analysis assumptions need to be made about the distributions of Y conditional on X (Y | X), X, and either U or W conditional on X (W | X). The latter specifications are equivalent under the assumptions of the simple additive model presented in the previous section. Reasonable choices for these distributions might be Y | X = x having a binary logistic regression with logit(p) = $\beta_0 + \beta_1 x$, X having a normal distribution with mean μ_x and variance σ_x^2, and W | X = x having a normal distribution with mean x and variance σ_w^2 (thus U has a normal distribution with mean 0 and variance σ_w^2). If these specifications can be made correctly there are advantages to using likelihood analysis. It is shown in Chapter 2, for example, that the likelihood ratio test is more reliable than other tests based

on approximate normality and standard errors, and that the efficiency of maximum likelihood estimates can be substantially important in some situations.

1.2.3 Assumptions for Likelihood Analysis

In usual regression models the explanatory variables are considered to be fixed, known constants. In regression models with measurement error in X the true explanatory variable, X, is unknown and the values must be considered either as fixed, unknown parameters (functional model) or as independent random variables from some probability distribution that depends on a fewer number of parameters (structural model). In order to do maximum likelihood analysis one generally assumes a structural model. If one were to assume a functional likelihood model then the unobserved true explanatory variable would be an unknown fixed quantity and thus be considered a nuisance parameter. This situation often results in as many nuisance parameters as cases in the data set, and maximum likelihood estimators in this situation may be very difficult to compute, are often not consistent, and generally not very useful.

Even using structural models, additional information about the measurement error is necessary to calculate useful maximum likelihood estimators. (This is a requirement of moment based methods as well.) This information can be in the form of a known distribution or known variance for the measurement error, or a subset of cases in the data set where the true explanatory variable is known or replicate measurements of the true explanatory variable.

1.2.4 Moment Based Methods

Moment based methods are a common approach for regression problems with explanatory variable measurement error. Fuller (1987) details moment methods for linear regression with measurement error. This approach usually involves making assumptions

about the first and second moments of the distributions in the measurement error problem rather than completely specifying them as in likelihood analysis. Sample moments are then equated with population moments in order to find estimators for the regression parameters. Caroll, Ruppert, and Stefanski (1995) present the "regression calibration" method for nonlinear regression models with explanatory variable measurement error. Here, a moment based estimate is found for E(X | W), and this estimate replaces X in the usual analysis that would be performed in the absence of measurement error.

1.2.5 A Simple but Historically Important Setting

A simple example which illustrates the problem of explanatory variable measurement error in regression is the attenuation problem in simple linear regression. Here we have a simple linear regression of Y on X, where it assumed the response variable, Y conditional on X = x has a normal distribution with mean $\beta_0 + \beta_1 x$ and variance σ_y^2 and X has a normal distribution with mean μ_x and variance σ_x^2. Instead of observing X one observes W, where W conditional on X = x is assumed to have normal distribution with mean x and variance σ_w^2 or equivalently, the measurement error, U is assumed to have a normal distribution with mean 0 and variance σ_w^2. In this situation the least squares estimate from the regression of Y on W, $\widehat{\beta}_1$, has an expected value equal to $\lambda\beta_1$, where

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_w^2} < 1.$$

Thus, measurement error in the explanatory variable has the effect of attenuating the regression line. An illustration of this effect can be seen in Figure 1.1 (in this example $\beta_0 = 0$, $\beta_1 = 1$, and $\sigma_y^2 = \sigma_x^2 = \sigma_w^2 = 1$, therefore $\lambda = 0.5$). It should be noted that in this situation under the assumption that σ_w^2 is known, Fuller's method-of-moments, "regression calibration", and maximum likelihood all result in the same estimator for β_1.

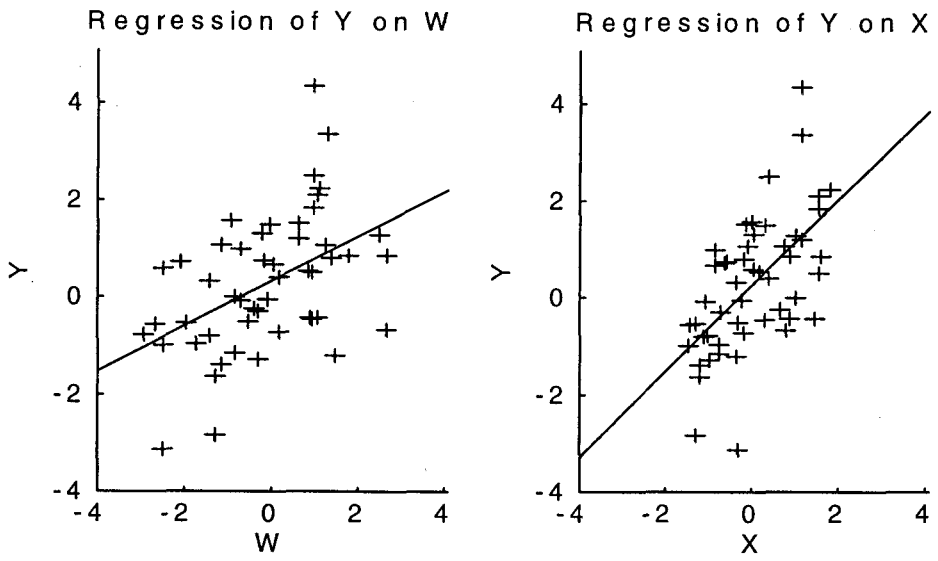

Figure 1.1. Attenuation in simple linear regression with explanatory variable measurement error.

1.3 Previous Studies

This section summarizes a number of previous related simulation studies in an attempt to indicate the current state of knowledge—from simulations—about efficiency, robustness, and inferential validity, particularly with respect to likelihood analysis. Schafer and Purdy (1996) examined the efficiency and inferential validity of maximum likelihood relative to moment based inference in the simple regression model where all of the underlying distributions are normal, when replicate measurements are available on a subset of cases. They showed maximum likelihood offered some gains in efficiency, but more importantly likelihood-ratio inference gave more valid tests and confidence intervals than the moment based methods. In their setting—in which all the distributions were normal—the likelihood and moment methods are nearly the same (they are the same if the measurement error variance is taken to be known). Thus, the efficiency differences

may be due to the incorporation of the information from replicate measurements in different ways; and the inferential validity comparisons may be due to that and the usual differences between likelihood-ratio and Wald inferences.

Schafer (1987) compared the mean squared errors of several estimators for the slope in a logistic regression with explanatory variable measurement error to an approximate maximum likelihood estimator. The approximate maximum likelihood estimator tended to have smaller mean squared error than the alternatives. An exact maximum likelihood estimator did not perform so well, but this was quite possibly due to convergence problems that were due to the choices involved in the numerical integral approximation.

Zhao and Lee (1997) examined the bias and efficiency of several estimators in logistic, Poisson and exponential-gamma regression models with explanatory variable measurement error. The study assumed the measurement error variance was known and distributions for the measurement error and true explanatory variable were both normal. Generally, an approximate likelihood estimate was as or more efficient in most of the simulated situations than the other estimators studied. The efficiency gains were most pronounced with large measurement error and smaller sample size.

Fuller (1987) showed in several simulations that bias corrections can reduce the mean square error for the maximum likelihood estimator in certain small sample situations in normal linear regression with known measurement error variance.

These studies, while not extensive, show the potential for efficiency gains for maximum likelihood estimators over moment based estimators and the advantage in reliability of likelihood ratio inference over the use asymptotic normality assumptions plus standard errors.

1.4 Overview

The main body of the thesis is divided into two main papers presented in manuscript format. The first, *Likelihood Analysis for Regression with Measurement*

Error, presents a general parametric model for measurement error in regression, a numerical algorithm to find maximum likelihood estimators and likelihood ratio statistics, and a simulation study to provide insight into the efficiency, validity and robustness of maximum likelihood analysis. The paper focuses primarily on the situations where a single explanatory variable is measured with error and replicate measurements of the explanatory variable are provided to give extra information about the nature of the measurement error. The second paper, *Maximum Likelihood Computations for Regression with Measurement Error*, focuses on the numerical methods for calculating maximum likelihood estimators and likelihood ratio statistics. This paper generalizes the numerical algorithm and computations of the first paper to include several other types of extra information about the measurement error and to incorporate product and quadratic terms involving the explanatory variable measured with error. These papers both use the EM algorithm (Dempster, Laird, and Rubin, 1977) for finding maximum likelihood estimators and use Gauss-Hermite quadrature for integral approximations.

2. LIKELIHOOD ANALYSIS FOR REGRESSION WITH MEASUREMENT ERROR

Roger Higdon and Daniel W. Schafer

2.1 Abstract

This paper advocates maximum likelihood analysis for generalized regression models with explanatory variable measurement error by (i) showing a computational algorithm for a wide variety of models, (ii) demonstrating the relative superiority of likelihood-ratio tests and confidence intervals over those based on asymptotic normality of estimates, and (iii) demonstrating that likelihood analysis is often more robust in these situations than has previously been feared. The ability to carry out likelihood analysis for a richer set of distributional assumptions than has been previously available, coupled with the encouraging robustness results, suggests that likelihood analysis may now have practical relevance for many regression problems with explanatory variable measurement error.

KEY WORDS: EM Algorithm; Errors-in-variables; Gauss-Hermite quadrature; Generalized linear models; Internal Replication; Nonlinear regression; Replicate measurements; Structural model; Surrogate variables.

2.2 Introduction

Likelihood analysis for regression with explanatory variable measurement error has received little attention relative to methods based on moment assumptions. This is due to computational difficulties, uncertainty about robustness, and the belief that methods based on weaker assumptions may perform just as well in practice. Now that computational tools are available for full likelihood analysis for a wide range of

measurement error models, it is important to reexamine its practical relevance. In particular, for those data problems that permit accurate distributional modeling, tests and confidence intervals based on likelihood ratios can be substantially more accurate than corresponding inferences based on approximate normality of commonly used estimators.

This paper provides a computational approach for full likelihood analysis for a broad range of structural models for linear and generalized linear models with explanatory variable measurement error. In addition it reports on the relative efficiency of maximum likelihood and common estimators when correct assumptions can be made, the validity of tests and confidence intervals based on both likelihood ratios and on estimates-plus-standard errors, and the robustness of likelihood inferences against departures from distributional assumptions.

The trade-offs between efficiency and test validity on the one hand, and robustness and transparency on the other hand are rather important in this setting because there are *several* distributional assumptions involved. Likelihood analysis, for example, requires the specification of three separate probability distributions. Thus there is more that can go wrong—more potential for model misspecification—but also more potential benefits in efficiency if the distributional assumptions are correctly made.

In addition, in many problems with explanatory variable measurement error there are different types of information from different cases in the data set. For example, there may be exact measurements or replicate measurements on a subset. In these settings there is some satisfaction that the likelihood methodology can appropriately and automatically incorporate the different types of information, and there is no need to carry out a two-stage analysis or to perform *ad hoc* weighting.

2.3 Model

The notation here follows that in Carroll, Ruppert, and Stefanski (1995). It is convenient to explicitly partition the explanatory variables into a set x that is only measured with error and a set z that is free of measurement error:

y is a univariate response

x is a vector of explanatory variables not directly observable

w is a measurement or surrogate for x, and

z is a vector of additional explanatory variables free of measurement error.

It is also convenient to partition z into possibly overlapping sets z_1, z_2, and z_3 to permit separate inclusion of explanatory variables in the three distinct parts of the structural regression-and-measurement error model. The density functions or probability mass functions for these are represented as follows:

$f(y \mid x, z_1; \theta_1)$: response distribution (1)

$f(w \mid x, z_2; \theta_2)$: measurement error distribution (2)

$f(x \mid z_3; \theta_3)$: unobserved explanatory variable distribution. (3)

The θ's are vectors of parameters and particular interest is in θ_1 or some subset of θ_1. The general formulation of (1) includes linear, generalized linear, and non-linear regression. For practical use we anticipate that (2) and (3) might be specified as normal linear models, generalized linear models, normal linear models on the log-scale, or mixtures of normal distributions. The mixture of normals for (2) and (3) has been suggested as a rich family of distributions that robustifies approaches based on full distributional assumptions, yet still leads to relatively easy likelihood calculations (Kuchenhoff and Carroll, 1997).

The parameter vector $\theta = (\theta_1, \theta_2, \theta_3)$ is not identifiable for many models and, even when it is, the analysis is generally impracticable without extra information (see, for example, Fuller, 1987, p. 9 and Carroll et al., p. 143). The following illustrate several possible types of "extra information:" *Situation 1—Known measurement error distribution:* Independent observations (y_i, w_i, z_i), for i = 1,...,n are available, arising from (1), (2), and (3); and θ_2 is known. *Situation 2—Internal validation:* Exact measurements x

are available for a subset of cases. That is, (y_i, w_i, x_i, z_i) are observed for i = 1,...,n_v; and (y_i, w_i, z_i) are observed for i = n_v + 1, ..., n_v + n_p. *Situation 3—Internal replication:* Replicate measurements are available for a subset of cases. That is, $(y_i, w_{i1}, \ldots, w_{ir_i}, z_i)$ are observed for i = 1,...,n; where r_i, the number of replicate measurements of x_i, is larger than 1 for at least some cases. *Situation 4—External validation:* Exact values and their measurements are available for a set of cases external to the primary data. So (x_i, w_i, z_i) are observed for cases in the validation data set, i = 1,...,n_v; and (y_i, w_i, z_i) are observed for cases in the primary data set, i = n_v + 1,..., n_v + n_p. *Situation 5—External replication:* Replicate measurements are available on a set of cases external to the primary data set. That is, $(w_{i1}, \ldots, w_{ir_i}, z_i)$ are observed for cases in the external replication set, i = 1,...,n_v; and (y_i, w_i, z_i) are observed in the primary data set, i = n_v + 1,..., n_v + n_p).

In all situations it is assumed that observations indexed by distinct values of i are mutually independent. It is also assumed that conditional on the true explanatory variable values, the measurements are independent of the responses:

$$f(y \mid x, w, z_1; \theta_1) = f(y \mid x, z_1; \theta_1). \qquad (4)$$

That is, the measurements should contain no additional information for predicting the response if the actual explanatory variables are available. This assumption is referred to as the "conditional independence assumption" or by Carroll et al. (1995) as "non-differential measurement error." It is a reasonable assumption for many data problems with measurement errors.

2.4 Likelihood Analysis

2.4.1 Maximum Likelihood and Moment Methods

A historically important model is the structural, simple linear regression model with normal distributions for (1) - (3). If $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, $x_i \sim N(\mu_x, \sigma_x^2)$, and $w_i \mid x_i$

~ $N(x_i, \sigma_w^2)$, and if the conditional independence assumption (non-differential measurement error) is true, then the joint distribution of (y_i, w_i) is bivariate normal. The minimal sufficient statistic from a sample of independent pairs (y_i, w_i) has dimension 5, so the 6 parameters are not identifiable. Maximum likelihood estimators may be obtained under the assumption that some parameter is known (see, for example, Madansky, 1959; Kendall and Stuart, 1979, Vol. 2, Chapter 29; and Fuller, 1987, p. 14). If σ_w^2 is known, for example, then the maximum likelihood estimator of β_1 is $S_{wy}/(S_{ww} - \sigma_w^2)$ (provided the denominator is positive), where S_{wy} is the sample covariance of w and y and S_{ww} is the sample variance of the w's. Since $E(S_{wy}) = \beta_1\sigma_x^2$, and $E(S_{ww}) = \sigma_x^2 + \sigma_w^2$, it is also a method of moments estimator. This estimator plays an important role as a starting point for many other method of moments-like estimators. Fuller (1987) has extended it—and more importantly, the multiple regression version of it—to many different types of linear models with various types of additional information.

It is also a special case of what has been labeled by Carroll et al. (1995, ch. 3) as the regression calibration approach, which uses the estimation technique that would have been used if x were available, but with x replaced by an estimate of E(x | w). See also Armstrong (1985). This starting point is important because method of moments and maximum likelihood coincide. More substantial differences between moment-based methods and maximum likelihood methods arise when one or more of the distributions are non-normal, when the extra information is in the form of internal replication or internal validation, and when the regression is nonlinear. One premise of this work is that the moment methods and their modifications do not tend to work very well in situations much different from linear regression with "everything normal". Although some believe that the moment methods are "robust" since they are based on weak assumptions, that belief does not seem to be supported in simulations. For example, the operating characteristics can be quite poor when the distribution of x is skewed.

Considerable attention has recently been given to situations in which (1) is specified by a generalized linear or nonlinear model (See Carroll et al., 1995, and references mentioned there). The regression calibration approach is currently quite popular because of its transparency: it employs the regression procedure that would be

used in the absence of measurement error, but with x replaced by an estimate of E(x | w). It should be noted, however, that the correct computation of standard errors can be difficult, and depends on the way in which E(x | w) is estimated (see Carroll et al., 1995, section 3.5). It is recognized that this easy approach is most appropriate when the measurement error is small since it usually involves a "small measurement error" approximation, i.e. a first-order expansion of E(y | x) about x = E(x | w). However, extensions have been made to allow for a quadratic approximation (see Carroll, Ruppert and Stefanski, 1995 ch. 3). Regardless of this potential shortcoming, more sophisticated techniques are not readily available at a practical level, nor have their relative merits been clarified.

2.4.2 Likelihood Analysis in Special Situations

Maximum likelihood estimation is based on the joint distribution of the observed random variables (y, w) (conditional on z). The likelihood must therefore be obtained from the model specified by (1)-(3) by integrating the joint distribution of (y, w, x) with respect to x.

Various researchers have investigated likelihood analysis by considering approximate techniques or techniques that bypass the integral, at least for particular distributional assumptions. Carroll, Spiegelman, Lan, Bailey and Abbot (1984) proposed a pseudo-likelihood technique for probit regression with normally-distributed measurement errors and normally-distributed x. In pseudo-likelihood, the nuisance parameters θ_2 and θ_3 in (2) and (3) are estimated in a first stage of the analysis and then treated as known in the likelihood function for θ_1. Schafer (1987) proposed an approach for approximate likelihood analysis for generalized linear models with normally-distributed measurement errors and normally-distributed x, using the EM algorithm. Crouch and Spiegelman (1990) suggested an approach for finding maximum likelihood estimators for logistic regression with normally-distributed measurement errors and normally-distributed x, using a particular integral approximation. Schafer (1993)

demonstrated likelihood analysis of probit regression with normally-distributed measurement errors and normally-distributed x, using the EM algorithm. Liu and Pierce (1994) suggested an approach for generalized linear models based on a Laplace approximation to the integral. Pseudo-likelihood was also investigated for a change point regression problem by Kuchenhoff and Carroll (1995). A Monte-Carlo computational technique for likelihood analysis was proposed by Kuha (1996).

We believe the current state of affairs regarding the practical use of likelihood analysis is the following: (i) Computational approaches have only been demonstrated for specialized settings so that "rich" modeling is not possible. (ii) Even though there are likely to be gains in efficiency by using likelihood analysis, there is an understandable concern about robustness, about which little is known. (iii) There are three probability distributions to be specified, and appropriate exploratory procedures for speculating on the three distributions are not automatically obvious. (iv) Even in the specialized settings, programs are not widely available, nor are they easy to write.

Nevertheless, we believe computational tools are now available for a broad class of models, that for many data problems the various submodels can be adequately specified (particularly when sample sizes are quite large), and that there is generally an under appreciation for the gains in efficiency and test validity from likelihood analysis when the models can be correctly specified. In the next section we illustrate one computational approach for full likelihood analysis.

2.5 Computational Issues

2.5.1 The Likelihood Function

For the remainder of this paper we shall focus on situation 3 (internal replication) and on x_i being a scalar. Treating the x_i's as random the likelihood function is $L(\theta) =$

$\prod_{i=1}^{n} f(y_i, \underline{w}_i | \underline{z}_i; \theta)$. Using the conditional independence assumption, this may be written in the form of the three model components as

$$L(\theta) = \prod_{i=1}^{n} \int f(y_i | x_i, \underline{z}_{i1}; \theta_1) f(\underline{w}_i | x_i, \underline{z}_{i2}; \theta_2) f(x_i | \underline{z}_{i3}; \theta_3) dx_i \qquad (5)$$

where $\underline{w}_i = (w_{i1}, ..., w_{ir_i})$.

Computational approaches for finding the parameter values that maximize the likelihood include those that attempt to evaluate the likelihood directly and those that bypass the likelihood with the EM algorithm. For the direct approach, the integral in (5) can be evaluated analytically if all distributions are normal (Carroll et al., 1995, Section 7.9.2; and Schafer and Purdy, 1996). More generally, it is necessary to embed a numerical approximation to the integral within a numerical optimization routine. Crouch and Spiegelman (1990) illustrated a trapezoidal rule-like approximation to the integral that may be used for binomial logistic regression with normally-distributed measurement errors and normally-distributed x. Carroll et al. (1995) found quadrature not to work so well in their experience. Liu and Pierce (1994) found the Laplace integral approximation to work well for approximating the likelihood in some situations but not others. Monte Carlo integration was investigated by McFadden (1989). In this, a pseudo-random sample from the distribution of $f(x | z_3; \theta_3)$ is generated and the integral is approximated as the average of $f(y | x, z_1; \theta_1) f(w | x, z_2; \theta_2)$ over the Monte Carlo distribution of x's.

2.5.2 EM Algorithm

If the x_i's were available then the likelihood function would be $L_c(\theta) = \prod_{i=1}^{n} f(y_i, \underline{w}_i, x_i | \underline{z}_i; \theta)$. Because the joint density in the product factors into three component models, the log of this likelihood can be conveniently expressed as $l_c(\theta) =$

$\sum_{i=1}^{n}[l_{1i}(\theta_1; y_i, x_i) + l_{2i}(\theta_2; \underline{w}_i, x_i) + l_{3i}(\theta_3; x_i)]$, where $l_{1i}(\theta_1; y_i) = \log f(y_i | x_i, \underline{z}_{i1}; \theta_1)$, $l_{2i}(\theta_2; \underline{w}_i, x_i) = \log f(\underline{w}_i | x_i, \underline{z}_{i2}; \theta_2)$, and $l_{3i}(\theta_3; x_i) = f(x_i | \underline{z}_{i3}; \theta_3)$.

The EM algorithm can be used to take advantage of this simple form by treating the x_i's as "missing data." If $\theta^{(t)}$ is the estimate after t iterations, the t+1 iteration of the algorithm requires an E-step: Compute

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} E[l_{1i}(\theta_1; y_i, x_i) + l_{2i}(\theta_2; \underline{w}_i, x_i) + l_{3i}(\theta_3; x_i) \mid y_i, \underline{w}_i; \theta^{(t)}]$$

and an M-step: chose $\theta^{(t+1)}$ to maximize $Q(\theta|\theta^{(t)})$.

The expectation in the M-step is with respect to the conditional distribution of x_i given y_i, $\underline{w}_i$, $\underline{z}_i$ and with unknown parameters in the expectation replaced by their current estimates. In general, then, $Q(\theta|\theta^{(t)}) =$

$$\sum_{i=1}^{n} \int [l_{1i}(\theta_1; y_i, x) + l_{2i}(\theta_2; \underline{w}_i, x) + l_{3i}(\theta_3; x)] \left\{ \frac{g_i(x)}{\int g_i(x^*)\, dx^*} \right\} dx \qquad (6)$$

where $g_i(x) = f(y_i | x, \underline{z}_{i1}; \theta_1^{(t)})\, f(\underline{w}_i | x, \underline{z}_{i2}; \theta_2^{(t)})\, f(x | \underline{z}_{i3}; \theta_3^{(t)})$.

The integrals can be evaluated exactly for the everything-normal linear regression model (Schafer and Purdy, 1996), for normal linear regression with normal mixture models for (2) or (3), and for binomial probit regression with normal measurement error and normal x (Schafer, 1994). A Laplace-like approximation was used for generalized linear models with normal measurement error and normal x by Schafer (1987). Kuha (1996) used Monte Carlo integration.

The approach here is to apply M-node Gauss-Hermite quadrature to both integrals. The result is that $Q(\theta|\theta^{(t)})$ in the E-step is replace by the approximation

$$Q^*(\theta|\theta^{(t)}) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{M} A_{ij}^{(t)}[l_{1i}(\theta_1;\, y_i,\, \tilde{x}_{ij}^{(t)}) + l_{2i}(\theta_2;\, \underline{w}_i,\, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3;\, \tilde{x}_{ij}^{(t)})] \qquad (7).$$

The development and details are provided below. First we shall describe how one uses (7).

Let $\theta^{(t)}$ be the "current" iterative value for the estimated parameter. Following a certain numerical maximization over x, one arrives at sampling values $\tilde{x}_{ij}^{(t)}$ for j from 1 to M (the number of nodes). The next iterative value for θ is the maximizer of (7) above. This can be maximized separately for each of the three terms, and each maximization amounts to simply a weighted analysis of the kind that would be used if the x_i's were available. In particular, one can often specify a generalized linear model for each of the component models, and then the derivatives for a Newton-Raphson Algorithm are simply weighted versions of the usual expressions.

The sampling points $\tilde{x}_{ij}^{(t)}$ and weights $A_{ij}^{(t)}$ in (7) are arrived at by the following argument. First consider the M-node Gauss-Hermite approximation to $\int g_i(x^*)\, dx^*$. In applying Gauss-Hermite quadrature, Liu and Pierce (1994) have pointed out the importance of transforming the variable of integration so that the sampling nodes represent values in an appropriate region for the integrand. Following their suggestion, let $\tilde{\mu}_i^{(t)}$ be the value that maximizes $g_i(x)$ and let $\tilde{\sigma}_i^{(t)}$ be $-\left[\partial^2 g_i(x)/\partial x^2\right]^2$ evaluated at $x = \tilde{\mu}_i^{(t)}$. Then transform the sampling points to $\tilde{x}_{ij}^{(t)} = \tilde{\mu}_i^{(t)} + \sqrt{2}\,\tilde{\sigma}_i^{(t)} u_j$, where u_j is the jth Gauss-Hermite quadrature node and use as the Gauss-Hermite approximation, $\sqrt{2}\,\tilde{\sigma}_i^{(t)} D_i^{(t)}$, where

$$D_i^{(t)} = \sum_{j=1}^{M} A_j \exp(u_j) f(y_i, \underline{w}_i, \tilde{x}_{ij}^{(t)} | \underline{z}_i;\, \theta^{(t)})$$

and where A_j is the weight attached to the jth node u_j (Abramowitz and Stegun, 1972, p. 924).

We then apply the idea to the numerator of (6) using the same transformation of the variable of integration. After some algebra the result is (7) above, where $A_{ij}^{(t)} = A_j \exp(u_j) f(y_i, \underline{w}_i, \tilde{x}_{ij}^{(t)} | \underline{z}_i; \theta^{(t)}) / D_i^{(t)}$. It is worth noting that M = 1 is equivalent to using the Laplace Approximation to the integrals. For the kinds of models in the next section we have found that an adequate value for M ranges from 1 or 2 nodes for logistic regression with the measurement error and x having normal distributions to 12 or more nodes for normal linear or logistic regression when x has a highly skewed gamma distribution and the measurement error variance is large. Finally, there are a few notes about the implementation and usage of this approach. First if x is strictly positive then it is appropriate to use a change of variables to log(x) before using quadrature. Second, although the maximized value of the log likelihood is not typically a by-product of the EM Algorithm calculations it does happen to be available in this instance. Notice that the denominator in (6) is the ith component in the product of (5). Based on the expressions above, therefore, the maximized value of the log likelihood is the value of $\sum_{i=1}^{n} \log(\sqrt{2}\,\tilde{\sigma}_i^{(t)} D_i^{(t)})$ at the final iteration. For likelihood ratio tests, for example, one can compute this maximized value for full and reduced models. To get confidence intervals for scalar parameters we consider a grid of possible values for the parameter then fit the reduced models at the grid points and retain in the confidence interval those values not rejected by the likelihood ratio at the appropriate level of significance.

2.6 Simulations

Figures 2.1-2.9 show the results of simulation studies comparing efficiency, validity, and robustness of various estimators under various structural models. The simple linear regression studies in Figures 2.1-2.4 are based loosely on the corn yield and soil nitrogen study presented by Fuller (1987, Secs. 1.2 and 3.1). In that example, about two-thirds of the observations had replicate measurements (of soil nitrogen). The simple logistic regression simulations in Figures 2.5-2.9 are base loosely on Clayton's (1991)

reported data on dietary saturated fat intake and heart disease mortality. In that case about one-fifth of the subjects had replicate measurements (of the saturated fat explanatory variable). The specific conditions examined are not too important and the details are relegated to the appendix (Section 2.8). Some consistent trends emerge, however, over the range of the different conditions examined.

The settings were such that in all cases the naive estimators—the ones that would be appropriate in the absence of measurement error—performed relatively poorly. These estimators were generally quite biased and thus coverage rates for confidence intervals based on them were inaccurate. Since the main focus here is on the relative performance of tools that *do* account for measurement error, results for the naive ones are not presented. The top panels for Figures 2.1-2.9 display the mean square errors of various estimators of the regression coefficient of x (over 1000 Monte Carlo samples) relative to the mean square errors of the maximum likelihood estimator or, in the cases of Figures 2.2, 2.3, 2.6, and 2.7 an approximate maximum likelihood estimator. The lower panels show the proportion of Monte Carlo samples for which the upper endpoint of a 95% confidence interval was less than the target value and the proportion for which the lower endpoint was greater than the target value. Ideally, these should be 2.5%. The reason for looking at the upper and lower error rates individually will soon be apparent. For maximum likelihood and approximate maximum likelihood estimators the confidence intervals are those based on likelihood ratios. For moment methods and the "regression calibration" method, they are based on asymptotic standard errors and approximate normality of estimators.

Each figure displays the results of simulations for one set of distributional assumptions, and "each simulation code" on a plot represents the results from 1000 simulated samples for one set of parameter values for that particular model. To convey a large number of results without too much clutter, the conditions are not labeled on the figures. Instead, the results are ordered from smallest to largest average relative mean squared error and assigned "simulation condition codes" accordingly. The codes mean different things in different figures. The parameter values corresponding to the codes are shown in the Appendix (Section 2.8).

Linear regression with skewed x. Figure 2.1 shows the results when y has a normal linear regression on x, the measurement error has a normal distribution, and the marginal distribution of x is gamma. The estimators considered here are the maximum likelihood estimator based on the actual conditions; the approximate maximum likelihood estimator based on the false assumption that x has a normal distribution; and the modified method-of-moments estimator suggested by Fuller (1987, Section 3.1). The most important feature, we believe, is the poor performance of confidence intervals based on the modified method-of-moments estimator. The explanation of this is provided in section 2.7. Next, it is important to note that the approximate maximum likelihood estimator based on an assumption of normality for x is not too bad here (Conditions 1, 3, 4, 6, 7, 8 10, and 15 have x from a gamma distribution with moderate skewness; the rest correspond to severe skewness). It is the comparison of the mean square error of this estimator to that of the modified method-of-moments, and similar comparisons that follow, that suggest to us thatthe method of moments may be less "robust" to the non-normality of x than is the "everything normal" maximum likelihood estimator. The relative MSE's tend to increase with increasing measurement error variance, increasing skewness in the distribution of x, and decreasing amount of cases with replicate measurements of x.

Linear regression with heavy-tailed x. Figure 2.2 reports similar comparisons when y has a normal linear regression on x, the measurement error is normal, and the marginal distribution of x is t with 3 degrees of freedom. In this case the computations of the previous section for obtaining maximum likelihood estimators are too difficult. The goal here is to examine the robustness of the estimators based on the incorrect assumption of normality for x, and the relative efficiency and robustness of a maximum likelihood estimator that uses a mixture of normal distributions for x. In this case the distribution of x was specified as a 95%/5% mixture of normal distributions with a common mean but different variances. Confidence intervals based on both of the approximate maximum likelihood estimators perform well here. With a large sample size (conditions 5, 6, 7, and 8) the maximum likelihood estimator that uses mixture of normals offers a sizable increase in efficiency over the other estimators.

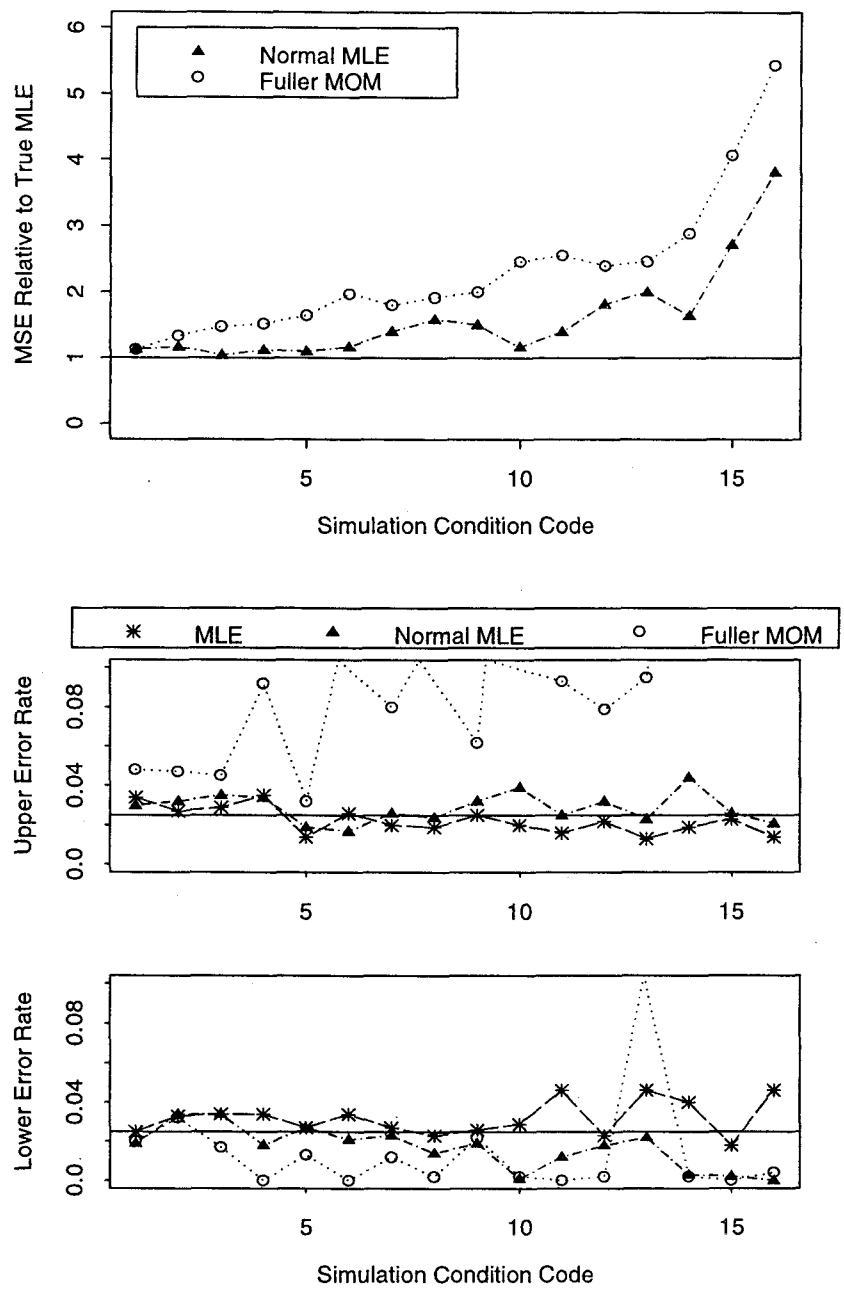

Figure 2.1. Plot of relative mean square errors and error rates for 95% confidence intervals for normal-normal-gamma model simulations. The Normal MLE is the one based on the (incorrect) normal-normal-normal model. See the appendix (Section 2.8) for parameter values corresponding to condition codes 1, 2, ... ,16.

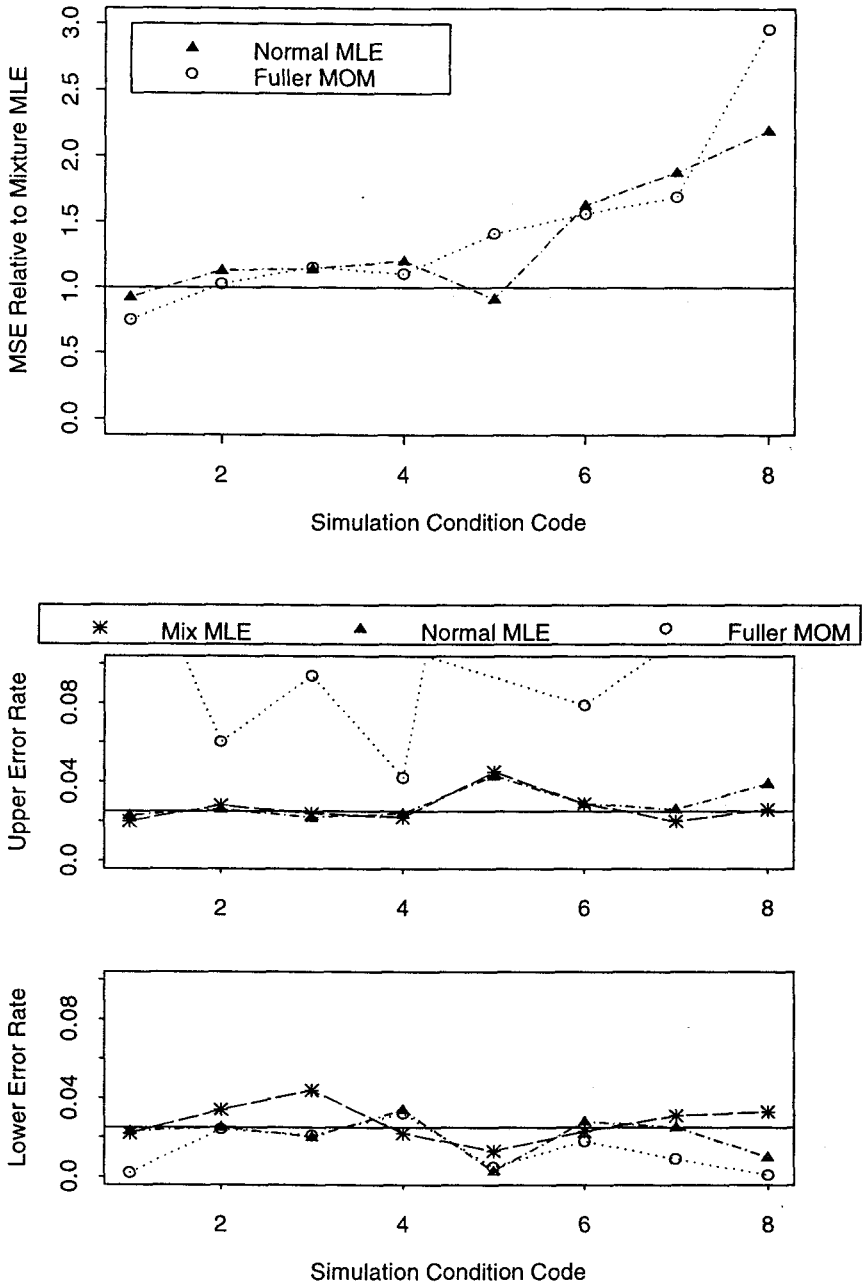

Figure 2.2. Plot of relative mean square errors and error rates for 95% confidence intervals for normal-normal-t model simulations.

Linear regression with heavy-tailed measurement error. In Figure 2.3 the measurement error has a t distribution with 3 degrees of freedom and x has a normal distribution. An approximate maximum likelihood estimator which uses a 95%/5% mixture of normal distributions with equal means but different variances for the measure error is used for comparison. It is quite apparent that the long tails of the measurement error distribution are a problem, since 95% confidence intervals based on the "everything normal" maximum likelihood estimator and the method-of-moments estimator are very inaccurate. The maximum likelihood estimator based on the mixture of normals also performs quite poorly, except in conditions where the sample size is large and the percentage of cases with replicate measurements is high (conditions 7 and 8). In those conditions the coverage rate of the 95% confidence intervals was reasonable and the mean squared error relative to the other estimators was much smaller. This suggests that a large amount of extra information—i.e. internal replication—is necessary to use the mixture of normals to model heavy-tailed measurement errors.

Linear regression with multiplicative measurement error. In Figure 2.4 a multiplicative measurement error model is used where both the measurement error and x have normal distributions on the log scale. For comparison to the exact maximum likelihood estimator and the "everything normal" maximum likelihood estimator an estimator due to Schafer (1992) is used. This estimator follows the "regression calibration" approach detailed by Carroll et al. (1995, Ch. 3). As stated in Section 2.4 the "regression calibration" estimator uses an estimate of $E(x \mid w)$ in place of x in the usual analysis. Here a quadratic approximation of $E(x \mid w)$ is used to improve the estimator because of the nonlinear relationship between x and w in the multiplicative model. The importance in accounting for the multiplicative nature of the measurement error is apparent in the conditions in which the log-normal distribution for x is more skewed (conditions 6, 8, 9, 10, 12, 14, 15, and 16). The "everything normal" maximum likelihood estimator does reasonably well, however, in the conditions when the distribution of x is less skewed. The quadratic "regression calibration" estimator performs quite poorly relative to the maximum likelihood estimator in most of the conditions.

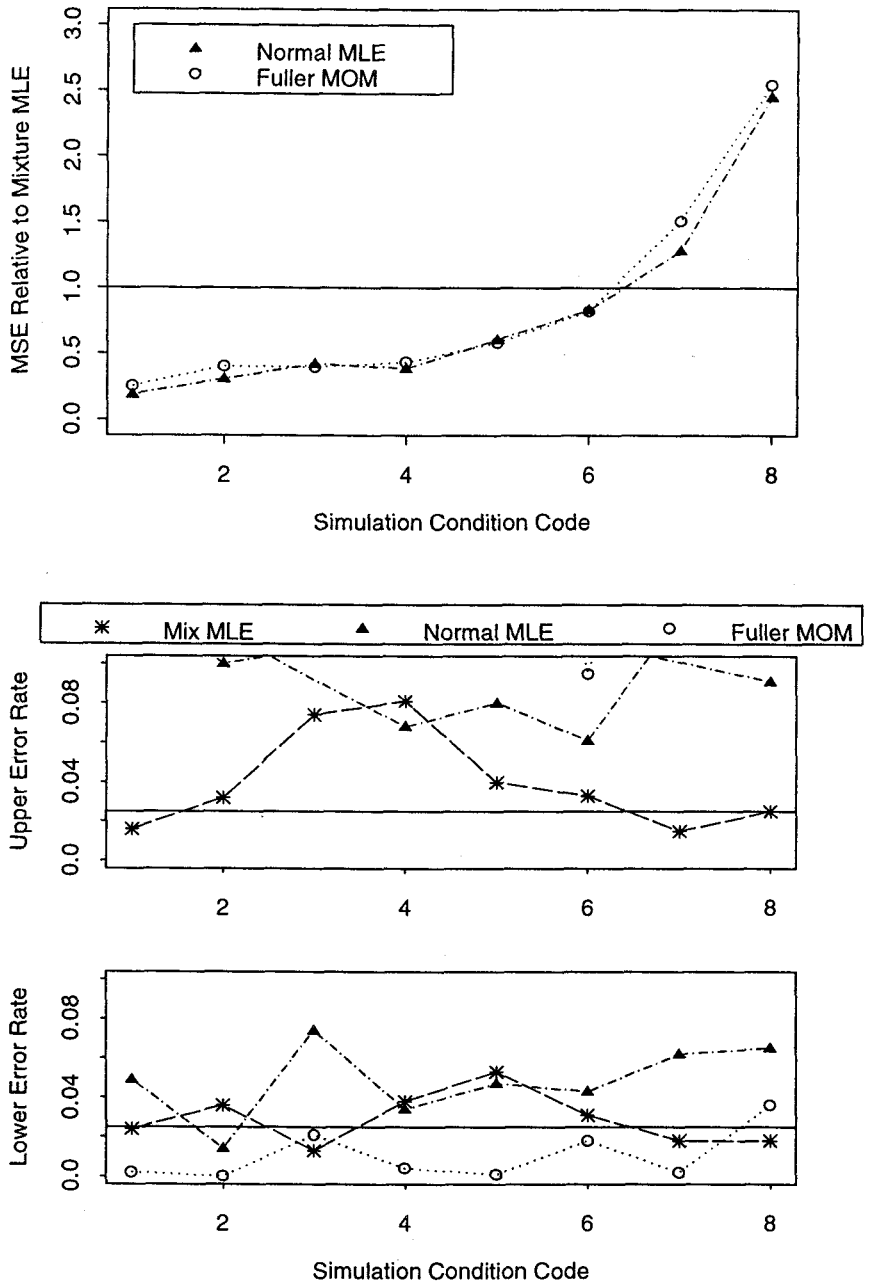

Figure 2.3. Plot of relative mean square errors and error rates for 95% confidence intervals for normal-t-normal model simulations.

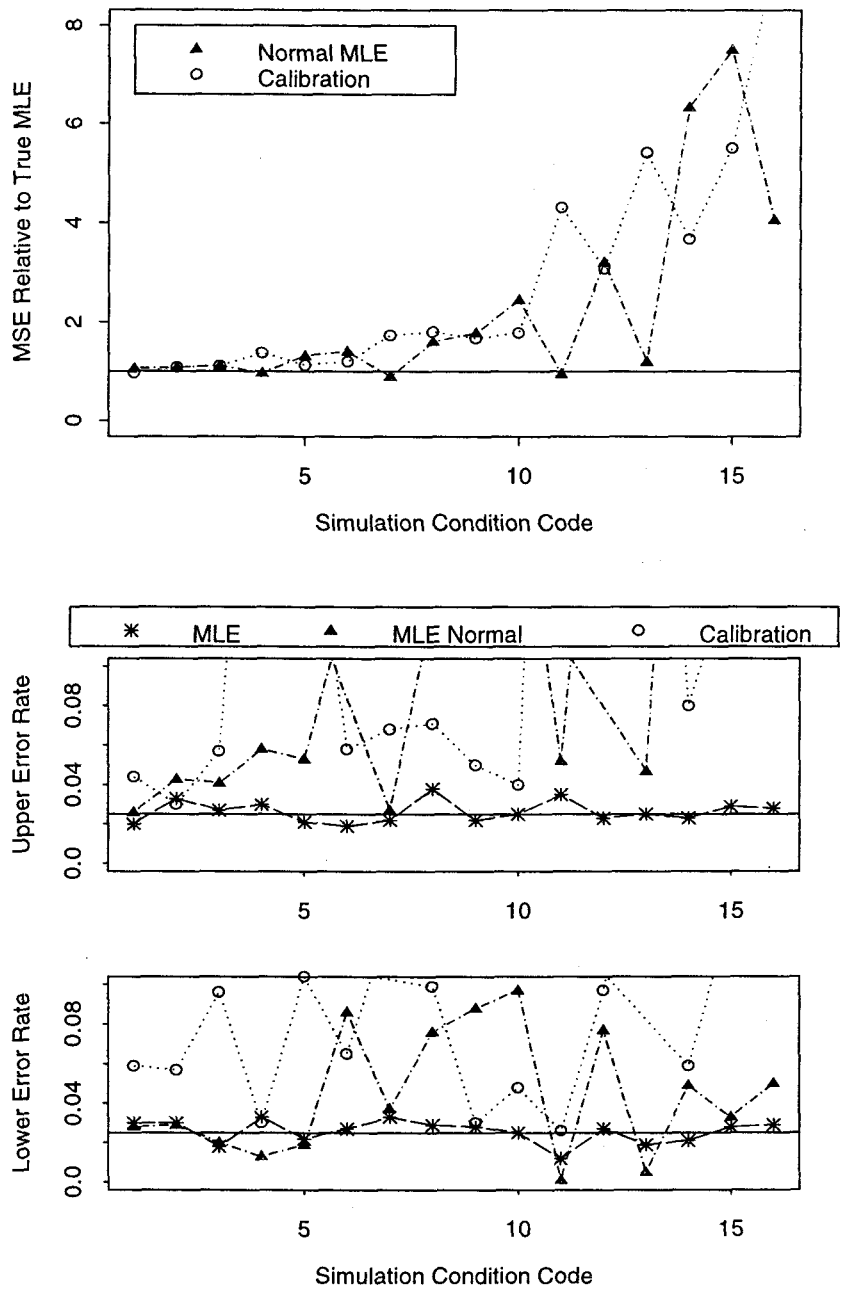

Figure 2.4. Plot of relative mean square errors and error rates for 95% confidence intervals for normal-lognormal-lognormal model simulations.

Binary logistic regression. In Figure 2.5, the measurement error distribution and the marginal distribution of x are both normal. In this case the maximum likelihood estimator is compared to the "regression calibration" estimator suggested by Carroll et al. (1995, Ch. 3). The relative efficiency of the two estimators is very similar except when the number of cases with replicate measurements is small and measurement error variance is large (conditions 7 and 8). In those cases the "regression calibration" confidence intervals based on approximate normality perform poorly. Carroll et al. (1995, Sec. A.3.3) present a likelihood ratio type test and its asymptotic distribution as an alternative to the standard error based test, but as can be seen in Figure 2.5 too few of the 95% confidence intervals based on it contain the true parameter value.

Binary logistic regression with skewed x. Figure 2.6 shows the situation when the measurement error is normal and x is gamma. The results here are quite consistent with those of Figures 2.1 and 2.5, with the relative MSE's increasing in the same conditions as in Figure 2.1 and the "regression calibration" estimator performing very poorly when the number of cases with replicate measurements is small and measurement error variance is large (conditions 13, 14, 15, and 16) . The comparisons here and those that follow also suggest, as was the case with the modified method-of-moments estimators, that the "regression calibration" is no more, and possibly less robust than the maximum likelihood estimator assuming that both the measurement error and x are normal—to departures from normality in the distribution of x.

Binary logistic regression with heavy-tailed x. In Figure 2.7 the measurement error is again normal, but the marginal distribution of x has a t distribution with 3 degrees of freedom. Here, as in Figure 2.1, mean squared error comparisons are made relative to an approximate maximum likelihood estimator based on x having a mixture of normal distributions. Here one notices that for the conditions with larger measurement error variance (conditions 5, 6, 7, and 8) the approximate maximum likelihood estimator based on a mixture distribution for x gives modest gains in efficiency over the other estimators. Confidence intervals based on the two approximate maximum likelihood estimators appear to be robust to this departure in model assumptions.

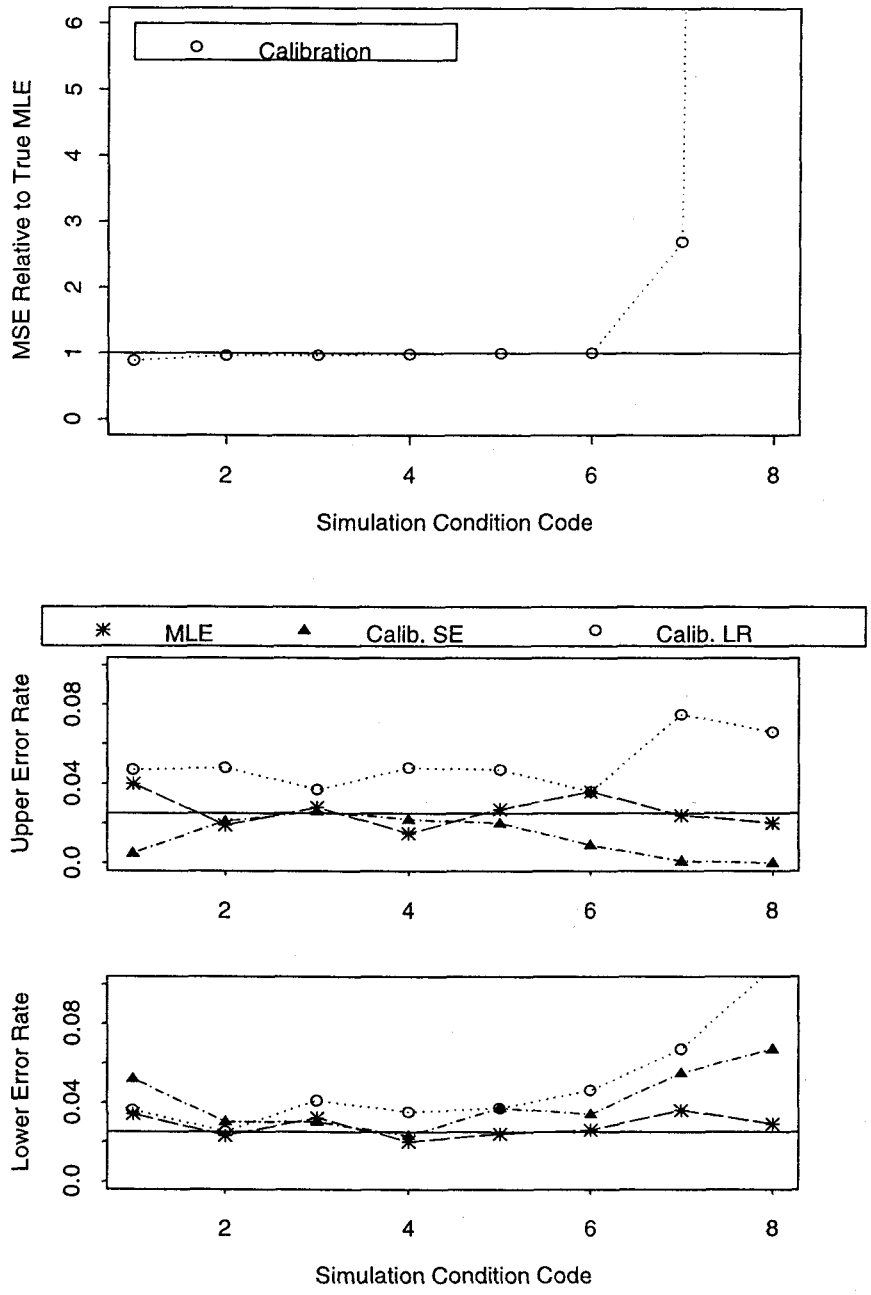

Figure 2.5. Plot of relative mean square errors and error rates for 95% confidence intervals for logistic-normal-normal model simulations. Calibr. SE and LR refer to confidence intervals based on the standard error and on a pseudo-likelihood ratio for the regression calibration estimator.

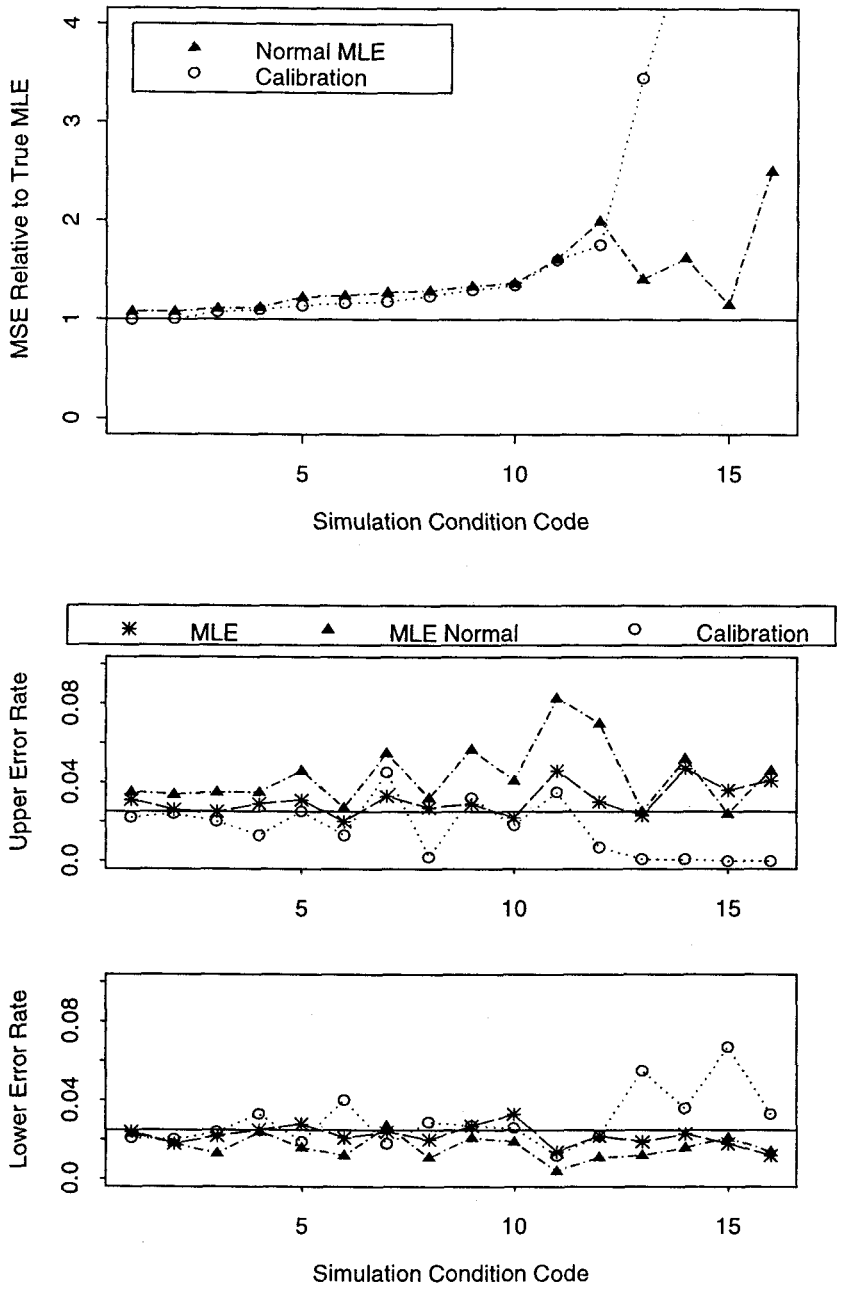

Figure 2.6. Plot of relative mean square errors and error rates for 95% confidence intervals for logistic-normal-gamma model simulations.

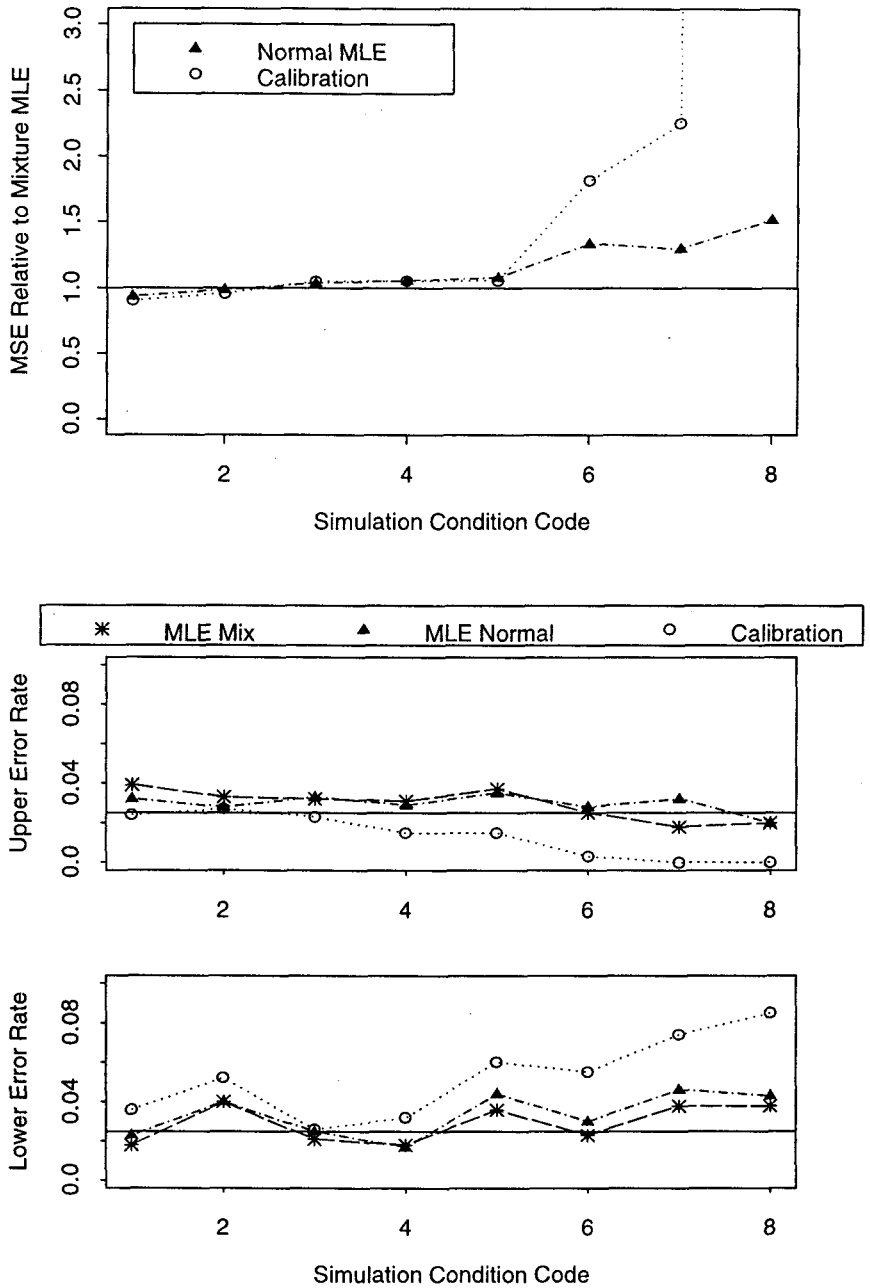

Figure 2.7. Plot of relative mean square errors and error rates for 95% confidence intervals for logistic-normal-t model simulations.

Binary logistic regression with heavy-tailed measurement error. In Figure 2.8 the measurement error has a t distribution with 3 degrees freedom and x has a normal distribution. Since an approximate maximum likelihood estimator which uses a mixture of normals for the measurement error is quite cumbersome to compute, the approximate maximum likelihood estimator which uses a mixture of normals for x is used as a basis for comparison instead. In conditions where the measurement error variance was large and the number of replicate measurements was not the largest (conditions 6, 7, and 8) none of the estimators were very robust to the departures from normality as indicated by the inaccuracy of the 95% confidence intervals. In the other conditions both of the approximate maximum likelihood estimators performed reasonably well. The "regression calibration" estimator, however, did not perform nearly as well.

Binary logistic regression with multiplicative measurement error. Figure 2.9 examines a multiplicative measurement error model. Both the measurement error and x have normal distributions on the log scale. For comparison to the maximum likelihood estimators a "regression calibration" estimator is used; however, here it is based on a multiplicative measurement error model and uses quadratic approximation for $E(x \mid w)$ as in Figure 2.4. The most important result here is the apparent necessity of using an estimator based on a multiplicative measurement error model. In conditions when the measurement error variance is larger (conditions 5, 6, 7, and 8) the approximate maximum likelihood estimator that assumes normal distributions for x and the measurement error is a good deal less efficient than the "true" maximum likelihood estimator and 95% confidence intervals based on it are inaccurate in these conditions. The "regression calibration" estimator does reasonably well in terms of mean square error, but suffers from the unreliability of asymptotic normality plus standard errors based confidence intervals.

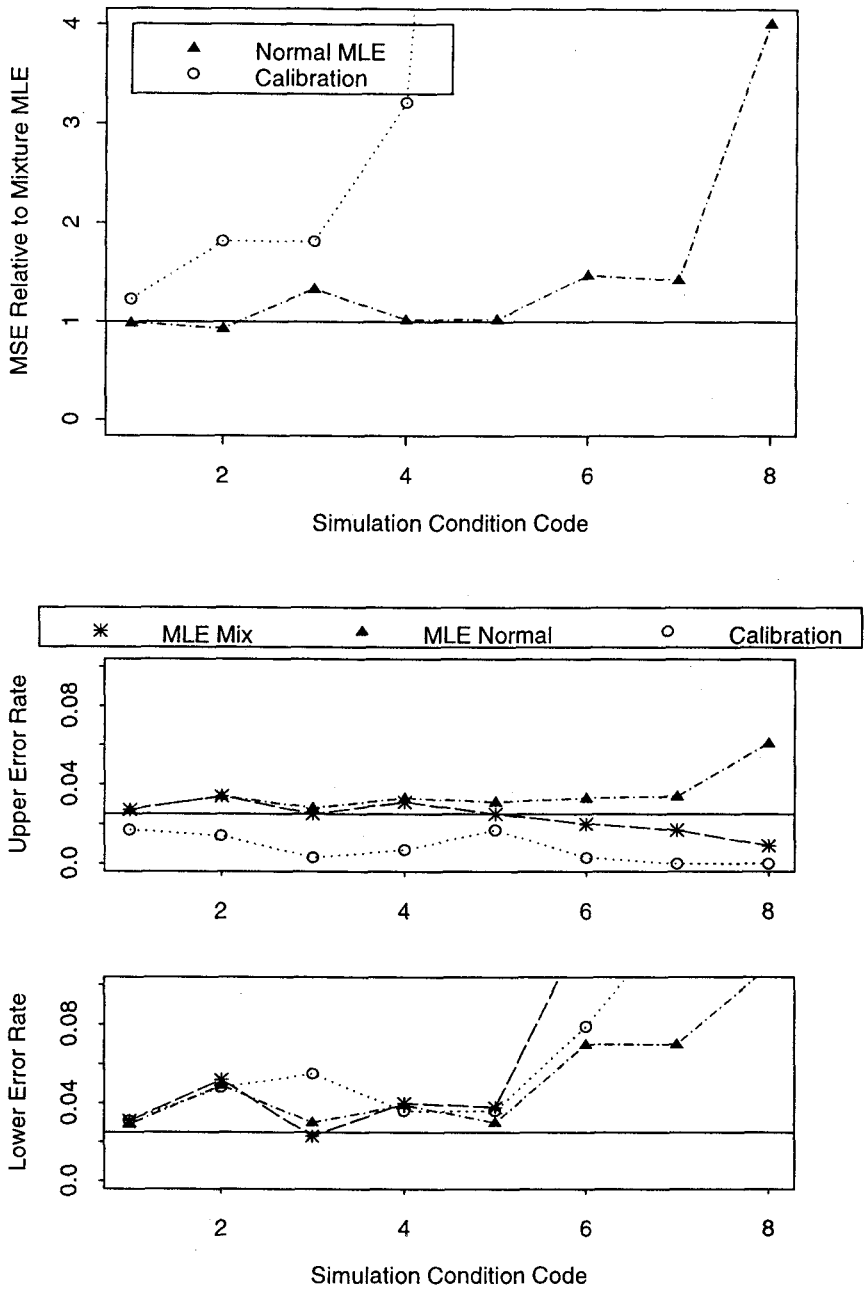

Figure 2.8. Plot of relative mean square errors and error rates for 95% confidence intervals for logistic-t-normal model simulations.

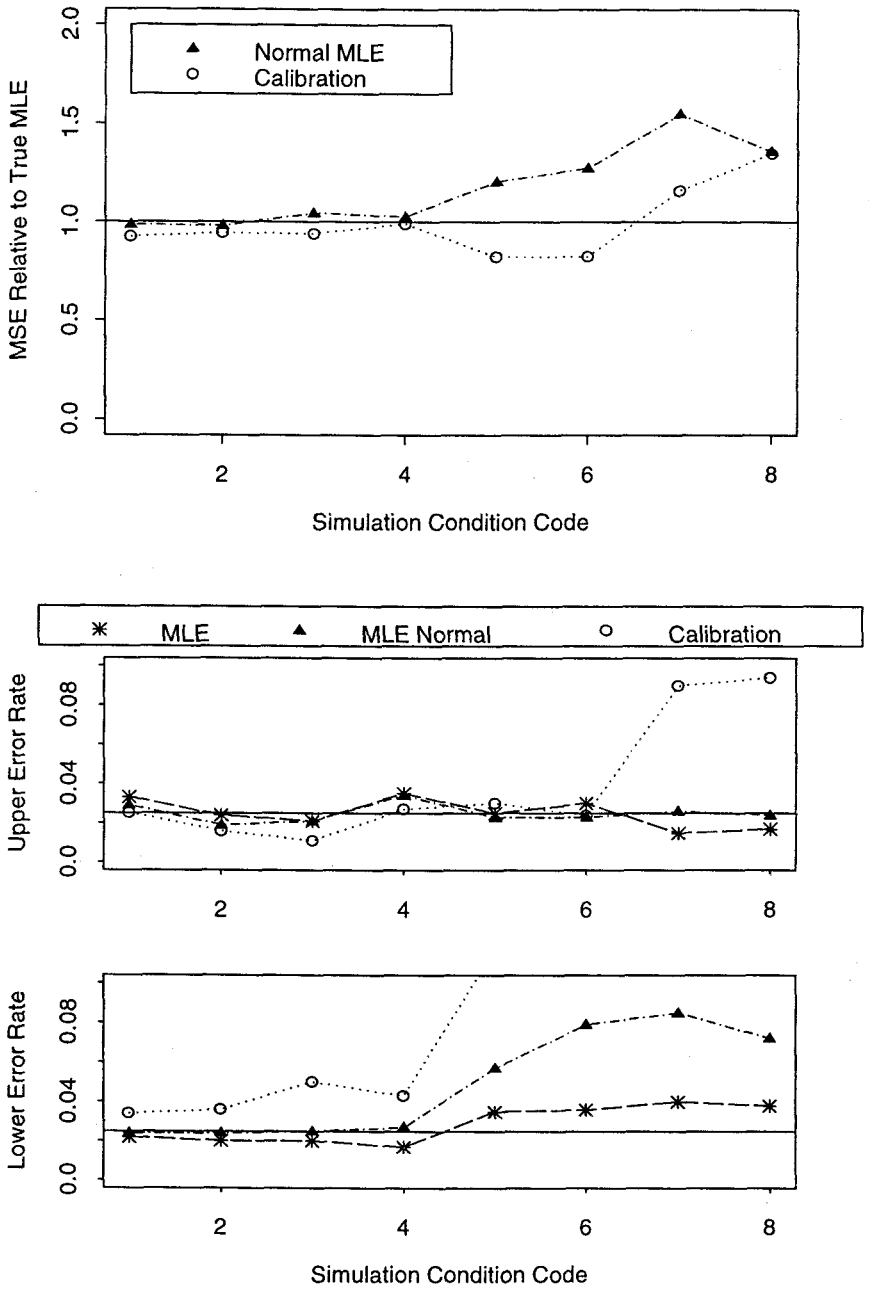

Figure 2.9. Plot of relative mean square errors and error rates for 95% confidence intervals for logistic-lognormal-lognormal model simulations.

2.7 Conclusions

The main point of this paper is that likelihood analysis is worth considering and its necessary computations are not out of the question. The computational tools and numerical comparisons in this paper have focused on the case that extra information is available in the form of internal replicate measurements. The computational extensions to other types of extra information are straightforward. The relative performance of estimators in other settings will not necessarily be the same, however. In fact, it might be expected that internal replication puts the most favorable light on likelihood analysis since other methods require some form of ad hoc weighting of cases based on different numbers of replicates. On the other hand, the consistent pattern of relative mean square errors, over different amounts of replication and different measurement error sizes is suggestive that likelihood analysis is at least worth considering in other situations as well.

2.7.1 Robustness of Likelihood Inference

The examinations into robustness of maximum likelihood inferences are necessarily limited in scope here. For example, we considered mostly one-at-a-time departures from the three assumed distributions. It did appear though that the "easy" maximum likelihood estimator—the one based on the "everything normal" model— is more robust than previously suspected. It appeared satisfactory when the distribution of x is skewed or heavy-tailed, but not so good when the distribution of the measurement error is heavy tailed or multiplicative in nature. Our thinking in reporting these studies is first that there is more flexibility in modeling the components to the structural measurement error model—for example by specifying the distribution of x (given z) to be gamma , lognormal or a mixture of normal distributions—and further that some degree of departure from these assumptions can be tolerated. The simulations of figures 2.3 and 2.8, however, suggest that when it is difficult to distinguish outliers as stemming from a heavy-tailed

distribution for x or a heavy-tailed distribution for measurement error, then subsequent analysis could be risky with any method.

2.7.2 Further Flexibility of Likelihood Analysis

It is important to realize there may be considerable advantage in specifying the distribution of x given additional explanatory variables z, rather than simply specifying the marginal distribution of x. With the computational technique proposed here, this may be accomplished with linear models, generalized linear models, and linear models after a log transformation. In addition, the incorporation of quadratic terms, x^2, and product terms, x∗z, is not difficult in this approach. In moment methods, by contrast, it would be necessary to specify the marginal distribution and the measurement error distribution (or at least the first and second moments of these) separately for x and for these terms constructed from x.

2.7.3 Poor Inferences Based on Asymptotic Standard Errors That Depend on $\widehat{\beta}$

The simulations demonstrate the poor performance of confidence intervals based on approximate normality of estimators and asymptotic standard errors. These intervals often have inaccurate coverage rates and are asymmetric in their coverage errors. That is the upper limit of a 95% confidence interval for the positive slope parameter is "too small" more than 2.5% of the time and the lower limit is "too large" less than 2.5% of the time. As illustrated in Figure 2.10, the standard error can depend quite heavily on the estimated coefficient of the explanatory variable measured with error. If there is a single explanatory variable with coefficient β then the standard deviation of the asymptotic distribution of $\widehat{\beta}$ typically has a piece involving $\beta \times$ measurement error variance. Using it requires the substitution of $\widehat{\beta}$ for β, but this results in standard errors that tend to be too small when $\widehat{\beta}$ is less than β and too large when $\widehat{\beta}$ is greater than β. As evident in the

simulation confidence intervals this effect can seriously affect the properties of the interval. The trouble, then, is in obtaining an estimate of asymptotic variance by substituting the estimate of β in place of the actual value. Thus, it should be noted the poor performance of confidence intervals of the moment based methods has nothing to do with the form of the extra information.

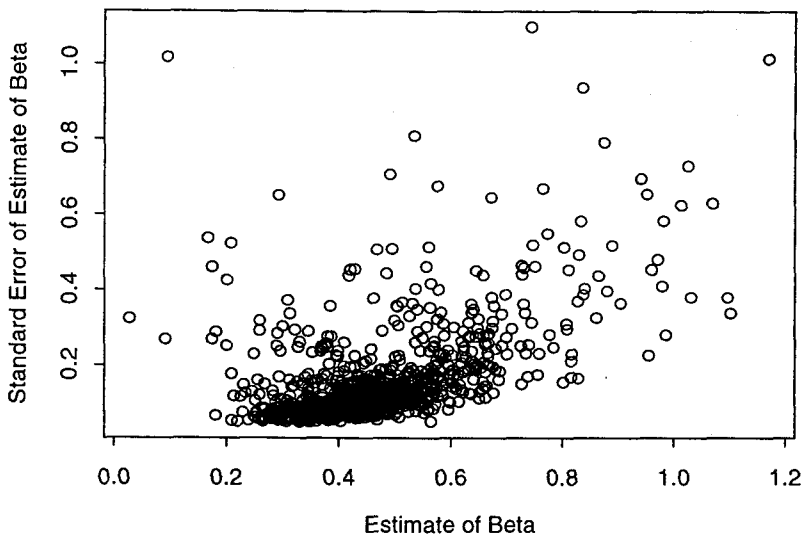

Figure 2.10. Plot of method-of-moments estimate of the slope and its standard error from the simulations described in Figure 2.1.

2.7.4 Performance of Moment Methods

The performance of moment based estimators were no better and often worse than the "everything normal" maximum likelihood estimator and generally much worse than the true maximum likelihood estimator when conditions deviated from the "everything normal" model. That the approach tends to work reasonably well for "everything normal" model is perhaps because that is the situation in which method-of-moments and maximum likelihood are very nearly the same.

2.7.5 Relevance of the EM Approach for Likelihood Calculations

It is not so much the purpose here to propose the EM algorithm with Gauss-Hermite quadrature as the "best" computation approach, but rather as one which is workable with current hardware limitations and which in some ways is fairly transparent. In theory, it extends easily to the case that more than one explanatory variable is measured with error, but the computation time when multiple integrals are approximated numerically (within the iterations of the EM algorithm) becomes a considerable hinderence. Kuha's (1996) approach of using Monte Carlo integration in the E-step may be more attractive in that case. Furthermore, if it is more convenient to specify the joint distribution of the imprecisely measured x's in terms of relevant conditional and marginal distribution, then Markov Chain Monte Carlo techniques could prove useful for that purpose. It is also not out of the question to throw the expression (5) with a numerical approximation directly into numerical non-linear optimization program. The advantage to this method would be in using pre-existing programs to calculate the maximum likelihood estimates. However, this method requires something equivalent to the E-step in the EM algorithm to approximate (5), and since the M-step is a relatively straightforward programming problem the advantages would not be great.

2.8 Appendix: Simulation Conditions

Figure 2.1 is based on 1000 Monte Carlo samples for each of the 16 conditions shown in the table below. The samples of y's and w's are generated from $y \mid x \sim N(60 + .5x, 60)$, $w \mid x \sim N(x, \sigma_w^2)$, and $x \sim$ Gamma(a, b) (shape parameter = a and scale parameter = b). There are 2 independent measurements w_{ij} for the first *rep* % of the n samples and single measurements for the remainder. The entries in the following table are the condition codes that are plotted on the x-axis of Figure 2.1.

Table 2.1. Condition Codes for Normal-Normal-Gamma Model.

		$\sigma_w^2 = 60$		$\sigma_w^2 = 240$	
(a,b)	n	rep = 20%	rep = 5%	rep = 20%	rep = 5%
4,8	100	3	4	6	10
	500	1	7	8	15
1,18	100	5	12	11	14
	500	2	9	13	16

The variance of x is 256 for the top 2 rows of the table and 324 for the bottom rows. The measurement reliabilities, $\sigma_x^2 / (\sigma_x^2 + \sigma_w^2)$ are therefore .81, .52, .84, and .57.

Figure 2.2 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x ~ N(60 + .5x, 60), w | x ~ N(x, σ_w^2), and x ~ $\sqrt{80} \cdot t_3 + 60$.

Table 2.2. Condition Codes for Normal-Normal-t Model.

	$\sigma_w^2 = 60$		$\sigma_w^2 = 240$	
n	rep = 20%	rep = 5%	rep = 20%	rep = 5%
100	2	3	1	5
500	4	6	7	8

The variance of x is 240 and therefore the measurement reliabilities are .8 and .5.

Figure 2.3 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x ~ N(60 + .5x, 60), w | x ~ $\sqrt{\sigma_w^2/3} \cdot t_3 + x$ and x ~ N(60, 240).

Table 2.3. Condition Codes for Normal-t-Normal Model.

	$\sigma_w^2 = 60$		$\sigma_w^2 = 240$	
n	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
100	*6*	*4*	*5*	*2*
500	*8*	*3*	*7*	*1*

The variance of x is 240 and therefore the measurement reliabilities are .8 and .5.

Figure 2.4 is based on 1000 Monte Carlo samples for each of the 16 conditions shown in the table below. The samples of y's and w's are generated from y | x ~ N(60 + .5x, 60), log(w) | log(x) ~ N(log(x), σ_w^2), and log(x) ~ N(4, σ_x^2) .

Table 2.4. Condition Codes for Normal-Logormal-Lognormal Model.

		$\sigma_w^2 = .01$		$\sigma_w^2 = .08$	
σ_x^2	*n*	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
.05	*100*	*1*	*7*	*4*	*11*
	500	*2*	*3*	*5*	*13*
.5	*100*	*6*	*8*	*12*	*16*
	500	*9*	*10*	*14*	*15*

The standard deviation of the measurement error is 10% of the value of x when σ_w^2 is .01 and 30% of value of x when σ_w^2 is .08.

Figure 2.5 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x having a binary logistic regression with logit(p) = 4 − 14x, w | x ~ N(x, σ_w^2), and x ~ N(4, .02).

Table 2.5. Condition Codes for Logistic-Normal-Normal Model.

	$\sigma_w^2 = .005$		$\sigma_w^2 = .02$	
n	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
300	*2*	*5*	*1*	*8*
1000	*3*	*4*	*6*	*7*

The variance of x is .02 and therefore the measurement reliabilities are .8 and .5.

Figure 2.6 is based on 1000 Monte Carlo samples for each of the 16 conditions shown in the table below. The samples of y's and w's are generated from y | x having a binary logistic regression with logit(p) = 2 − .05x, w | x ~ N(x, σ_w^2), and x ~ Gamma(a, b) (shape parameter = a and scale parameter = b).

Table 2.6. Condition Codes for Logistic-Normal-Gamma Model.

		$\sigma_w^2 = 36$		$\sigma_w^2 = 144$	
(a,b)	*n*	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
4,6	*300*	*1*	*4*	*8*	*15*
	1000	*2*	*3*	*6*	*13*
1,12	*300*	*5*	*10*	*12*	*16*
	1000	*7*	*9*	*11*	*14*

The variance of x is 144 and therefore the measurement reliabilities are .8 and .5.

Figure 2.7 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x having a binary logistic regression with logit(p) = 14 − 4x, w | x ~ N(x, σ_w^2), and x ~ $\sqrt{.00667} \cdot t_3 + 4$.

Table 2.7. Condition Codes for Logistic-Normal-t Model.

	$\sigma_w^2 = .005$		$\sigma_w^2 = .02$	
n	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
300	*1*	*4*	*6*	*8*
1000	*2*	*3*	*5*	*7*

The variance of x is .02 and therefore the measurement reliabilities are .8 and .5.

Figure 2.8 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x having a binary logistic regression with logit(p) = 14 − 4x, w | x ~ $\sqrt{\sigma_w^2/3} \cdot t_3$ + x and x ~ N(4, .02).

Table 2.8. Condition Codes for Logistic-t-Normal Model.

	$\sigma_w^2 = .005$		$\sigma_w^2 = .02$	
n	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
300	*1*	*2*	*6*	*7*
1000	*5*	*4*	*3*	*8*

The variance of x is .02 and therefore the measurement reliabilities are .8 and .5.

Figure 2.9 is based on 1000 Monte Carlo samples for each of the 8 conditions shown in the table below. The samples of y's and w's are generated from y | x having a binary logistic regression with logit(p) = 14 − 4x, log(w) | log(x) ~ N(log(x), σ_w^2) and x ~ N(4, .1).

Table 2.9. Condition Codes for Logistic-Lognormal-Lognormal Model.

	$\sigma_w^2 = .01$		$\sigma_w^2 = .08$	
n	*rep = 20%*	*rep = 5%*	*rep = 20%*	*rep = 5%*
300	*1*	*4*	*4*	*8*
1000	*2*	*3*	*6*	*7*

The standard deviation of the measurement error is 10% of the value of x when σ_w^2 is .01 and 30% of value of x when σ_w^2 is .08.

2.9 References

Abramowitz, M., and Stegun J. (1972), *Handbook of Mathematical functions*, New York: Dover.

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

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, New York: Chapman and Hall.

Carroll, R. J., Spiegelman, C., Lan, K. K., and Abbott, R. D. (1984), "On Errors-in-Variables for Binary Regression Models," *Biometrika*, 71, 19-26.

Clayton, D. G. (1991), "Models for the Analysis of Cohort and Case-control Studies with Inaccurately Measured Data," In Dwyer, J. H., Feinleib, M., Lippert, P., and Hoffmeister, H. (Editors), *Statistical Models for Longitudinal Studies*, New York: Oxford University Press, 301-331.

Crouch E. A. and Spiegelman D. (1990), "The Evaluation of Integrals of the Form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)dt$: Applications to Logistic-Normal Models," *Journal of the American Statistical Association*, 85, 464-467.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

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

Kendall, M. G., and Stuart, A. (1979) *The Advanced Theory of Statistics, Vol.* 2, London: Griffin.

Kuchenoff, H., and Carroll, R. J. (1997), "Biases in Segmented Regression with Errors in Predictors," *Statistics in Medicine*, 439-450.

Kuha, J. (1996), "Monte Carlo Methods for Adjusting for Measurement Error," *The XVIIth International Biometric Conference.*

Liu, Q., and Pierce, D. A. (1994), "A Note on Gauss-Hermite Quadrature," *Biometrika*, 81, 624-629.

Liu, X., and Liang, K. Y. (1992), "Efficacy of Repeated Measures in Regression Models with Measurement Error," *Biometrics*, 48, 645-654.

Mandansky, A. (1959), "The Fitting of Straight Lines when Both Variables are Subject to Error," *Journal of the American Statistical Association*, 54, 173-205.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 29-265.

Schafer, D. W. (1987), "Covariate Measurement Error in Generalized Linear Models," *Biometrika*, 74, 385-391.

Schafer, D. W. (1992), "Replacement Methods for Measurement Error Models," *Oregon State University Technical Report 151.*

Schafer, D. W. (1993), "Likelihood Analysis for Probit Regression with Measurement Errors," *Biometrika*, 80, 899-904.

Schafer, D. W., and Purdy, K. G. (1996), "Likelihood Analysis for Errors-in-Variables Regression with Replicate Measurements," *Biometrika*, 83, 813-824

Stefanski, L. A., and Carroll, R. J. (1985) "Covariate Measurement Error in Logistic Regression," *Annals of Statistics*, 8, 407-422.

3. MAXIMUM LIKELIHOOD COMPUTATIONS FOR REGRESSION WITH MEASUREMENT ERROR

Roger Higdon and Daniel W. Schafer

3.1 Abstract

This paper will present a general computational method for maximum likelihood analysis for generalized regression with measurement error in a single explanatory variable. The method will use the EM algorithm in conjunction with Gauss-Hermite quadrature in the E-step. This method will allow the use of maximum likelihood estimation under a fairly broad range of distributional assumptions, and thus will allow one to gain the efficiency and inferential advantages of likelihood analysis, principally, the ability to use likelihood ratio inference instead of less reliable asymptotic normality plus standard errors methods.

Keywords: Measurement error, EM algorithm, Gauss-Hermite quadrature, Generalized linear models, Structural model

3.2 Introduction

Many different methods have been used to find estimates in linear and generalized linear models with explanatory variable measurement error. Particular attention has been given to moment based methods (see Fuller, 1987, and Carroll, Ruppert and Stefanski, 1995). On the other hand, much less attention has been given to likelihood based methods, possibly due to the computational complexity and the difficulty in checking parametric assumptions. There have been uses of likelihood estimation in regression models with measurement error, such as Schafer (1987 and 1992), Schafer and Purdy

(1996) and Kuha (1996). Some recent work, Schafer and Purdy (1996), Higdon and Schafer (1998) has shown advantages of likelihood methods over other methods in certain situations, particularly in regard to hypothesis tests and confidence intervals using the likelihood ratio statistic. Therefore, it seems important to develop general methods and computer programs to perform maximum likelihood estimation in regression problems with explanatory variable measurement error.

This paper provides a general computation method for likelihood analysis in a wide range of structural models for linear and generalized linear models with a single explanatory variable measured with error. This computational method uses the EM algorithm of Dempster, Laird and Rubin (1977) along with modified Gauss-Hermite quadrature (Pierce and Liu, 1993) to approximate integrals in the E-step. The details of this approach were laid out by Higdon and Schafer (1998) for the special case that extra information about the measurement error distribution is available in the form of internal replicate measurements on a subset of observations. This paper presents the computational details for other types of extra information, including known measurement error distribution, internal validation data, external validation data, and external replication.

3.3 Model

3.3.1 Notation

The notation for this model will follow that of Carroll et. al. (1995) for structural models. The explanatory variables will be partitioned into two sets: the first, a scalar explanatory variable x that is measured with error, and the second a vector of explanatory variables z measured without error.

y: a univariate response

x: a univariate explanatory variable not directly observable for some or all the data points

w: a measurement or surrogate for x (with possibly several w's for each x)

z: a vector of additional explanatory variables with measurement error

It is also convenient to partition z into possibly overlapping sets z_1, z_2, and z_3 to allow for the separate inclusion of explanatory variables in the three distinct components of the structural model. The densities or probability mass functions which make up the likelihood for this model are represented as follows.

$f(y \mid x, z_1; \theta_1)$: response distribution (1)

$f(w \mid x, z_2; \theta_2)$: measurement error distribution (2)

$f(x \mid z_3; \theta_3)$: unobserved explanatory variable distribution (3)

The θ's are vectors of parameters and θ_1 or some subset of θ_1 is of primary interest. The general formulation of (1) in practice will include linear, generalized linear, and non-linear regression models . For practical use we anticipate that (2) and (3) might be specified as normal linear models, generalized linear models, normal linear models on the log-scale, or mixtures of normal distributions. A mixture of normals for (2) and (3) has been suggested as a rich family of distributions that robustifies approaches based on full distributional assumptions, yet still leads to relatively easy likelihood calculations (Kuchenhoff and Carroll, 1995).

3.3.2 Extra Information

The parameter vector $\theta = (\theta_1, \theta_2, \theta_3)$ is not identifiable for many models, such as when (1) is a normal linear regression and (2) and (3) are chosen to have normal distributions (See Fuller, 1987 p. 9). Even when θ is identifiable, as in the case when (1) is a binary logistic regression and (2) and (3) have normal distributions, there is little practical information contained about the parameters without the inclusion of extra information in the data set (See Carroll et. al., 1995, p.143). The following illustrate several possible types of "extra information:"

Situation 1: Known measurement error distribution. Independent observations (y_i, w_i, z_i), for i = 1,...,n are available, arising from (1), (2), and (3); and θ_2 is known. This is rarely a realistic situation. But if there is no additional information about the measurement error distribution, one might conduct an analysis with θ_2 presumed known and explore the sensitivity of the results to the particular choice.

Situation 2: Internal validation. Exact measurements x are available for a subset of cases. That is, (y_i, w_i, x_i, z_i) are observed for i = 1,...,n_v; and (y_i, w_i, z_i) are observed for i = n_v + 1, ..., $n_v + n_p$.

Situation 3: Internal replication. Replicate measurements are available for a subset of cases. That is, $(y_i, w_{i1}, \ldots, w_{ir_i}, z_i)$ are observed for i = 1,...,n; where r_i, the number of replicate measurements of x_i, is larger than 1 for at least some cases .

Situation 4: External validation. Exact values and their measurements are available for a set of cases external to the primary data. So (x_i, w_i, z_i) are observed for cases in the validation data set, i = 1,...,n_1; and (y_i, w_i, z_i) are observed for cases in the primary data set, i = n_1 + 1,..., n.

Situation 5: External replication. Replicate measurements are available on a set of cases external to the primary data set. That is, $(w_{i1}, \ldots, w_{ir_i}, z_i)$ are observed for cases in the external replication set, $i = 1,\ldots,n_1$; and (y_i, w_i, z_i) are observed in the primary data set, $i = n_1 + 1,\ldots, n$.

3.3.3 Further Assumptions

It is assumed that observations indexed by different i's are mutually independent. It is also assumed that conditional on the true explanatory variable values, the measurements are independent of the responses:

$$f(y \mid x, w, z_1; \theta_1) = f(y \mid x, z_1; \theta_1).$$

That is, the measurements should contain no additional information for predicting the response if the actual explanatory variables are available. This assumption is referred to as the "conditional independence assumption" or by Carroll et. al. (1995) as "non-differential measurement error." It is a reasonable assumption for many data problems with measurement errors.

3.4 The Likelihood

The structural model likelihood combines the models (1), (2), and (3), and incorporates the different types of "extra information" presented in the previous section. Two different likelihoods will be defined; the complete data likelihood, $L_c(\theta)$, and the observed data likelihood, $L_o(\theta)$. The complete data likelihood would be the likelihood if there were a complete set of (y_i, w_i, x_i, z_i) for each case internal to the data set and a complete set of (w_i, x_i, z_i) for any cases external to the data set as in situations 4 and 5. The observed data likelihood is obtained by integrating over all of the unobserved x's in

the complete data likelihood. The goal of maximum likelihood estimation will be to maximize the observed data likelihood. The complete data likelihood for situations 1,2 and 3 is

$$L_c(\theta) = \prod_{i=1}^{n} f(y_i|\, x_i, z_{i1};\, \theta_1)\, f(w_{i1},\ldots,w_{ir_i}|\, x_i, z_{i2};\, \theta_2)\, f(x_i|\, z_{i3};\, \theta_3)$$

where, n is the total number of cases in the data set and r_i is the number of replicate measurements of x_i. Also, r_i may be 1 for all cases in situations 1 (known measurement error) and 2 (internal validation). In many situations one may assume that $w_{i1},\ldots,w_{ir_i}$ are independent after conditioning on x_i, in that case $f(w_{i1},\ldots,w_{ir_i}|\, x_i, z_{i2};\, \theta_2) = \prod_{j=1}^{r_i} f(w_{ij}|\, x_i, z_{i2};\, \theta_2)$. This need not be the case, however. For situations 4 and 5 the complete data likelihood is

$$L_c(\theta) = \prod_{i=1}^{n_1} f(w_{i1},\ldots,w_{ir_i}|\, x_i, z_{i2};\, \theta_2)\, f(x_i|\, z_{i3};\, \theta_3) \times \prod_{i=n_1+1}^{n} f(y_i|\, x_i, z_{i1};\, \theta_1)\, f(w_{i1},\ldots,w_{ir_i}|\, x_i, z_{i2};\, \theta_2)\, f(x_i|\, z_{i3};\, \theta_3)$$

where, n_1 is the total number of external cases in the data set and n is the total number of cases internal and external in the data set, and r_i is equal to 1 for situation 4. Integrating over the cases with unobserved x_i's will give the observed data likelihoods:

(Situations 1 and 3)

$$L_o(\theta) = \prod_{i=1}^{n} \int f(y_i|\, x_i, z_{i1};\, \theta_1)\, f(w_{i1},\ldots,w_{ir_i}|\, x_i, z_{i2};\, \theta_2)\, f(x_i|\, z_{i3};\, \theta_3)\, dx_i$$

where, for situation 1, θ_2 is assumed known and $r_i = 1$.

(Situation 2)

$$L_o(\theta) = \prod_{i=1}^{n_1} f(y_i|\, x_i, z_{i1}; \theta_1) f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3) \times \prod_{i=n_1+1}^{n} \int f(y_i|\, x_i, z_{i1}; \theta_1)\, f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3)\, dx_i$$

where, n_1 is the total number of cases with observed x_i's.

(Situation 4)

$$L_o(\theta) = \prod_{i=1}^{n_1} f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3) \times \prod_{i=n_1+1}^{n} \int f(y_i|\, x_i, z_{i1}; \theta_1)\, f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3)\, dx_i.$$

(Situation 5)

$$L_o(\theta;) = \prod_{i=1}^{n_1} \int f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3)\, dx_i \times \prod_{i=n_1+1}^{n} \int f(y_i|\, x_i, z_{i1}; \theta_1)\, f(w_i|\, x_i, z_{i2}; \theta_2)\, f(x_i|\, z_{i3}; \theta_3)\, dx_i.$$

It may also be possible to have combinations of the above situations occurring in the same data set and in those cases the likelihoods can be adjusted accordingly.

3.5 Maximizing the Likelihood Through the EM Algorithm

Maximum likelihood estimators will be found by maximizing the observed data log-likelihood $l_o = \log L_o$. The approach presented here to maximize the observed data log-likelihood is to treat the true explanatory variable x as missing and use the EM algorithm. One could try to maximize the observed data likelihood directly by evaluating

or approximating the integrals of the previous section and plugging the results into a non-linear optimization program. As it turns out the programming details of this approach are nearly as involved as using the EM algorithm. Also, many of these optimization programs are a good deal slower at finding the maximum likelihood estimates than is the EM algorithm for this problem. The EM algorithm also has some unique benefits: the maximized value of the likelihood (for likelihood ratio inferences) is a by-product of the computations in this setting and asymptotic standard errors can be computed with the approach of Louis (1982) .

3.5.1 The EM algorithm

The EM (expectation-maximization) algorithm of Dempster et. al. (1977) is an iterative procedure designed to provide maximum likelihood estimators in the presence of missing data. Each iteration of the EM algorithm has two steps; the E-step (expectation), where one computes the expectation of the complete data log-likelihood, l_c = Log L_c given the observed data and the M-step (maximization), where one maximizes the quantity computed in the E-step with respect to the parameter vector θ. Dempster et. al. (1977) showed that the value of the observed data likelihood evaluated at this estimate is increased with each successive iteration.

An initial estimate, $\theta^{(0)}$of the parameter is required to begin the algorithm. At each successive iteration this estimate is updated, so that at the end of the t-th iteration the estimate of θ is $\theta^{(t)}$. In the E-step of the t+1-th iteration one needs to compute the quantity $Q(\theta|\theta^{(t)})$, which is the expected value of the complete data log-likelihood given the observed data and with θ replaced $\theta^{(t)}$ in the conditional distribution. In terms of the notation for the measurement error problem and situations 1 and 3 this gives

$$Q(\theta|\theta^{(t)}) = E\{l_c(\theta) \mid y, w; \theta^{(t)}\}.$$

$Q(\theta|\theta^{(t)})$ is then maximized with respect to θ in the M-step, giving $\theta^{(t+1)}$, the updated estimate of θ. The algorithm is continued until a given level of convergence is achieved. That is until $|\theta_j^{(t+1)} - \theta_j^{(t)}|/|\theta_j^{(t+1)}| < \epsilon$, for each element θ_j of θ and the given level of convergence ϵ.

Applying the EM algorithm to regression with explanatory variable measurement error involves three separate problems, (i) finding an initial estimate of the parameter vector θ, (ii) formulating $Q(\theta|\theta^{(t)})$ in the E-step, which involves evaluating certain integrals, and (iii) maximizing $Q(\theta|\theta^{(t)})$ in the M-step, which involves the maximization of several weighted log-likelihoods.

3.5.2 The E-step

The E-step involves formulating the quantity $Q(\theta|\theta^{(t)})$. In regression with measurement error and internal replication (situation 3) this formulation becomes

$$Q(\theta|\theta^{(t)}) = E\{l_c(\theta) \mid y, w; \theta^{(t)}\}$$

$$= \sum_{i=1}^{n} \int l_{ci}(\theta) \frac{f(y_i, w_{i1},..., w_{ir_i}, x_i \mid z_i; \theta^{(t)})}{\int f(y_i, w_{i1},..., w_{ir_i}, x_i^* \mid z_i; \theta^{(t)})\, dx_i^*} dx_i \qquad (4)$$

where $f(y_i, w_{i1},..., w_{ir_i}, x_i \mid z_i; \theta)$ may be specified in terms of the three densities in (1)-(3) (assuming the conditional independence assumption) as

$$f(y_i, w_{i1},..., w_{ir_i}, x_i \mid z_i; \theta) = f(y_i \mid x_i, z_{i1}; \theta_1)\, f(w_{i1},...,w_{ir_i} \mid x_i, z_{i2}; \theta_2)\, f(x_i \mid z_{i3}; \theta_3).$$

Also, $l_{ci}(\theta)$ is the contribution to the complete data log likelihood from case i and may be written as $l_{1i}(\theta_1; y_i, x_i) + l_{2i}(\theta_2; w_{i1},..., w_{ir_i}, x_i) + l_{3i}(\theta_3; x_i)$, where $l_{1i}(\theta_1; y_i, x_i) = \log f(y_i \mid x_i, z_{i1}; \theta_1)$, $l_{2i}(\theta_2; w_{i1},..., w_{ir_i}, x_i) = \log f(w_{i1},...,w_{ir_i} \mid x_i, z_{i2}; \theta_2)$, and $l_{3i}(\theta_3; x_i) = \log f(x_i \mid z_{i3}; \theta_3)$. Additionally, if we assume that $w_{i1},...,w_{ir_i}$ are independent after conditioning

on x_i, then $f(w_{i1},...,w_{ir_i} | x_i, z_{i2}; \theta_2) = \prod_{j=1}^{r_i} f(w_{ij} | x_i, z_{i2}; \theta_2)$ and $l_{2i}(\theta_2; w_{i1},..., w_{ir_i}, x_i) = \sum_{j=1}^{r_i} \log f(w_{ij} | x_i, z_{i2}; \theta_2)$.

The principal problem of the E-step becomes evaluating the integrals in (4). In some situations these integrals can be evaluated exactly, such as in the "everything normal" model which assumes normal distribution for (1) - (3) (See Schafer and Purdy, 1996) or in the case where y (1) has a probit regression and both the measurement error and x are taken to have normal distributions (See Schafer, 1993). Also, in the case where y (1) has a normal linear regression and either the measurement error (2) or x (3) is taken to be a mixture of normal distributions these integrals can be calculated exactly. However, in other cases these integrals will have to be approximated numerically.

3.5.2.1 Gauss-Hermite Quadrature

The approach here will be to approximate the integral using Gauss-Hermite quadrature. Gaussian quadrature approximates an integral by taking a weighted sum of the integrand evaluated at several sampling points. These points (nodes) are the roots of orthogonal polynomials. The specific orthogonal polynomials used depends on the limits of integration; for example, Hermite polynomials are used if the limits of integration are from $-\infty$ to ∞. If M nodes are used the weights are chosen so that the approximation of the integral of a polynomial of degree 2M times a weighting function is exact. The weighting function for Gauss-Hermite quadrature is a normal density. Therefore, if a function can be approximated well by a polynomial of degree 2M times a normal density then Gauss-Hermite quadrature will work well. The weights and nodes for various type of quadrature are given in Abramowitz and Stegun (1972) for several different numbers of nodes.

Since Gauss-Hermite quadrature is intended for functions defined over the entire real line, if the variable of integration, x, is strictly positive a change of variable to log(x) is appropriate. Blindly applying Gauss-Hermite quadrature by directly using the nodes

and weights given in Abramowitz and Stegun (1972) will often give very poor results. It is very important to transform the variable of integration so that the integrand be sampled in an appropriate region. One can show quite simply how to re-express Gauss-Hermite quadrature by transforming the original nodes u_i to $\tilde{x}_i = \mu + \sqrt{2} \cdot \sigma u_i$ (Naylor and Smith, 1982). The problem now becomes choosing an appropriate μ and σ. Following Liu and Pierce (1994) if one wishes to integrate f(x), then choose μ to be the mode of f(x) and $\sigma = 1/\sqrt{j}$, where $j = -\frac{\partial^2}{\partial x^2} f(x)\Big|_{x=\mu}$. Thus, $\int f(x)dx \approx \sqrt{2} \cdot \sigma \sum_{i=1}^{M} A_i \exp(u_i) f(\tilde{x}_i)$, where A_i, is the weight attached to the ith node. Applying this to the denominator in (4), let $\tilde{\mu}_i^{(t)}$ be the value of x_i that maximizes $f(y_i, w_{i1},..., w_{ir_i}, x_i \mid z_i; \theta^{(t)})$ and let $\tilde{\sigma}_i^{(t)}$ be the square root of the reciprocal of $-l''_{ci}(\theta^{(t)})\Big|_{x_i = \tilde{\mu}_i^{(t)}}$ evaluated at this mode. The Gauss-Hermite sampling nodes for the integral are then taken to be $\tilde{x}_{ij}^{(t)} = \tilde{\mu}_i^{(t)} + \sqrt{2} \cdot \tilde{\sigma}_i^{(t)} u_j$ for j = 1,...,M (for M-node quadrature); where the u_j's are the zeros of the Mth order Hermite polynomial. The integral in the denominator of (4), which is the contribution of case i to the observed data likelihood is approximated by $\sqrt{2} \cdot \tilde{\sigma}_i^{(t)} D_i^{(t)}$, where

$$D_i^{(t)} = \sum_{j=1}^{M} A_j \exp(u_j) f(y_i, w_{i1},..., w_{ir_i}, \tilde{x}_{ij}^{(t)} \mid z_i; \theta^{(t)})$$

where A_j is the weight attached to the jth node u_j (Abramowitz and Stegun, 1972, p. 924). It should be noted with this shifting and rescaling, one-node Gauss-Hermite quadrature is equivalent to the Laplace approximation (Liu and Pierce, 1994).

For the integral in the numerator of (4) it is reasonable to use the same transformation of the variable of integration as in the denominator since, f(x) log f(x) tends to zero as f(x) does and the mass of the function is centered at the same point as as f(x). So, this means that

$$Q(\theta \mid \theta^{(t)}) \approx$$

$$\sum_{i=1}^{n} \sum_{j=1}^{M} A_{ij}^{(t)} [l_{1i}(\theta_1; y_i, \tilde{x}_{ij}^{(t)}) + l_{2i}(\theta_2; w_{i1},.., w_{ir_i}, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3; \tilde{x}_{ij}^{(t)})] \quad (5)$$

where $A_{ij}^{(t)} = A_j \exp(u_j) f(y_i, w_{i1},..., w_{ir_i}, \tilde{x}_{ij}^{(t)} | z_i; \theta^{(t)})/D_i^{(t)}$.

3.5.2.2 Adjustments to Q for Different Extra Information

The adjustments to (5) for different types of extra information are straightforward. Situation 1 (known measurement error) is simply a special case of situation 3, where θ_2 is known and $r_i = 1$ for all i. For situations where there is external data (situations 4 and 5) $l_{ci}(\theta)$ will not contain $l_{1i}(\theta_1; y_i, x_i)$ for the external data. When there is validation data (situations 2 and 4) $l_{ci}(\theta) = l_{oi}(\theta)$ for those cases with validation data. Therefore, there is no need to evaluate an integral and for these cases $l_{ci}(\theta)$ can be added directly to (5) with weight of 1. For the different types of extra information the approximations for $Q(\theta|\theta^{(t)})$ are as follows:

Situation 2 (internal validation)

$$Q(\theta|\theta^{(t)}) \approx \sum_{i=1}^{n_1}[l_{1i}(\theta_1; y_i, x_i) + l_{2i}(\theta_2; w_i, x_i) + l_{3i}(\theta_3; x_i)] +$$

$$\sum_{i=n_1+1}^{n}\sum_{j=1}^{M} A_{ij}^{(t)}[l_{1i}(\theta_1; y_i, \tilde{x}_{ij}^{(t)}) + l_{2i}(\theta_2; w_i, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3; \tilde{x}_{ij}^{(t)})]$$

where, n_i is the number of cases with observed x_i's.

Situation 4 (external validation)

$$Q(\theta|\theta^{(t)}) \approx \sum_{i=1}^{n_1}[l_{2i}(\theta_2; w_i, x_i) + l_{3i}(\theta_3; x_i)] +$$

$$\sum_{i=n_1+1}^{n}\sum_{j=1}^{M} A_{ij}^{(t)}[l_{1i}(\theta_1; y_i, \tilde{x}_{ij}^{(t)}) + l_{2i}(\theta_2; w_i, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3; \tilde{x}_{ij}^{(t)})]$$

where, n_1 is the number of external cases.

Situation 5 (external replication)

$$Q(\theta|\theta^{(t)}) \approx \sum_{i=1}^{n_1}\sum_{j=1}^{M} A_{ij}^{(t)}[\, l_{2i}(\theta_2;\, w_{i1},..,\, w_{ir_i},\, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3;\, \tilde{x}_{ij}^{(t)})] +$$

$$\sum_{i=n_1+1}^{n}\sum_{j=1}^{M} A_{ij}^{(t)}[l_{1i}(\theta_1;\, y_i,\, \tilde{x}_{ij}^{(t)}) + l_{2i}(\theta_2;\, w_i,\, \tilde{x}_{ij}^{(t)}) + l_{3i}(\theta_3;\, \tilde{x}_{ij}^{(t)})]$$

where, n_1 is the number of external cases.

3.5.2.3 Finding the Modes for Quadrature Transformations

One of the main computational obstacles to the Gauss-Hermite quadrature approach is determining the $\tilde{\mu}_i^{(t)}$'s, the modes with respect to x of the components of the complete data likelihoods at the t-th iteration. In some cases these modes can be solved for directly, as is the case when y is has a normal linear regression, the measurement error has a normal distribution, and x has a gamma distribution. In most situations, however these modes need to be found numerically. The problem here is that n functions need to be maximized simultaneously. A straightforward approach is to use a few iterations of the Newton-Raphson Algorithm to approximate the modes. A good starting value is often necessary for the Newton-Raphson Algorithm to work well. A reasonable starting value for the modes on the first iteration of the algorithm would be $\tilde{\mu}_i^{(0)} = \frac{1}{2}(\overline{w}_i + \hat{\mu}_x^{(0)})$ (or $\log\, \tilde{\mu}_i^{(0)}$ in cases when a log transformation of x is used for the integral), where $\overline{w}_i = \sum_{j=1}^{r_i} w_{ij}/r_i$ and $\hat{\mu}_x^{(0)}$ is the mean of the $\overline{w}_i$'s or perhaps x_i's when there is validation data. Then, for the t+1-th iteration use the $\tilde{\mu}_i^{(t)}$'s.

3.5.2.4 Choosing the Number of Nodes for Quadrature

Another issue in the E-step is choosing the number of nodes to be used in Gauss-Hermite quadrature. Abramowitz and Stegun (1972) provide tables with up to 20 nodes. The tradeoff here is between accuracy of the integrals in the E-step and computational complexity. Therefore, the goal should be to achieve a desired level of accuracy with as few nodes as possible. To provide some insight and guidelines as to the number of nodes required for different circumstances the results of a small simulation study are provided in Table 3.1. The simulation conditions are identical to those used in Higdon and Schafer

Table 3.1. Accuracy Relative to 20 Node Gauss-Hermite Quadrature for Slope Estimates.

Best Case

Model	*2 nodes*	*6 nodes*	*12 nodes*
Normal-Normal-Gamma	*.1325*	*.0032*	5.66×10^{-4}
Normal-Lognormal-Lognormal	*.0026*	4.04×10^{-7}	4.17×10^{-12}
Logistic-Normal-Normal	8.25×10^{-4}	2.06×10^{-9}	6.38×10^{-12}
Logistic-Normal-Gamma	*.1454*	*.0022*	4.75×10^{-4}
Logistic-Lognormal-Lognormal	5.06×10^{-4}	6.05×10^{-8}	1.09×10^{-11}

Worst Case

Model	*2 nodes*	*6 nodes*	*12 nodes*
Normal-Normal-Gamma	*2.737*	*.5861*	*.0367*
Normal-Lognormal-Lognormal	*.1398*	*.0057*	*.0011*
Logistic-Normal-Normal	*.0755*	3.65×10^{-4}	3.99×10^{-7}
Logistic-Normal-Gamma	***	***	*.0665*
Logistic-Lognormal-Lognormal	*.0508*	*.0030*	1.43×10^{-4}

*** *Because of the inaccuracy of the quadrature, singularity problems arose in the numerical procedures.*

(1998). Those simulations examined maximum likelihood estimation for simple linear and logistic regression under several combinations of conditions and distributions for the measurement error and for x. The simulations varied the sample size, the percentage of

cases with replicate measurements, the size of the measurement error variance, and the degree of skewness in the distribution of x. Here the simulations compare the average accuracy of estimates of the slope based on 2, 6, 12, and 20 node Gauss-Hermite quadrature for 50 Monte-Carlo samples. The table presents the relative accuracy of each of the estimators for the condition in which they were most accurate and least accurate for each of the distributional combinations, (1) - (3) in the structural model. The relative accuracy is defined as the mean of $|\widehat{\beta}_M - \widehat{\beta}_{20}| / \widehat{\beta}_{20}$ for the 50 Monte-Carlo samples, where $\widehat{\beta}_M$ is estimate of the slope based on M quadrature nodes. In general each of the estimators is more accurate when the measurement error was smaller and when the distribution of x is less skewed. The sample size and percentage of cases with replicate measurements has much less impact.

3.5.3 The M-step

Compared to the numerical difficulties in the E-step the M-step is quite straightforward. The M-step consists of maximizing $Q(\theta|\theta^{(t)})$ with respect to θ to obtain the updated estimate $\theta^{(t+1)}$. The form of the expression for $Q(\theta|\theta^{(t)})$ (5) obtained in the E-step is that of three separate weighted log-likelihoods, which allows maximization for the three component parameter vectors θ_1, θ_2, and θ_3 separately. The calculations in the M-step will not be difficult if the distributions in (1) - (3) are ones for which maximum likelihood analysis are routinely used. For example, in the case of normal linear models, exact expressions for the updated parameter estimates can be found, while for generalized linear models, like logistic and poisson regression, updating the estimates from $\theta^{(t)}$ to $\theta^{(t+1)}$ involves familiar Newton-Raphson calculations which can be done with pre-existing software routines or simple programs.

3.5.4 Calculating the Maximized Observed Data Log-likelihood and Standard Errors

This approach to maximizing the observed data likelihood allows for easy calculation of the maximized observed data log-likelihood, which can be used for likelihood ratio tests and confidence intervals based on inverting the test. Since the integral in the denominator of (4) is the contribution of the ith data point to the observed data likelihood, the calculation of the maximized value of the observed log-likelihood, $l_o(\widehat{\theta})$, simply requires the computation of one more E-step after the last iteration, t = T. The maximized value for situations 1, 3, and 5 is

$$l_o(\widehat{\theta}) \approx \sum_{i=1}^{n} \log(\sqrt{2}\,\tilde{\sigma}_i^{(T)} \mathrm{D}_i^{(T)})$$

where $\widehat{\theta}$ is approximated by $\theta^{(T)}$. The addition of terms to the sum for cases with validation data can be done without approximation, thus for situations 2 and 4 the maximized value of the observed log-likelihood is

$$l_o(\widehat{\theta}) \approx \sum_{i=1}^{n_1} l_{ci}(\theta^{(T)}) + \sum_{i=n_1+1}^{n} \log(\sqrt{2}\,\tilde{\sigma}_i^{(T)} \mathrm{D}_i^{(T)}).$$

This expression can be calculated for a full and reduced model in order to compute a likelihood ratio test statistic. To get confidence intervals for scalar parameters we consider a grid of possible values for the parameter then fit the reduced models at the grid points and retain in the confidence interval those values not rejected by the likelihood ratio test at the appropriate level of significance.

Although likelihood ratio inference is generally more reliable than inference based on asymptotic normality plus standard errors, one can obtain asymptotic standard error estimates for Wald type inference and confidence intervals through a method due to Louis (1982). This method uses the result that the observed information matrix can be represented by

$$\frac{-\partial^2 l_o(\theta)}{\partial\theta^2} = -\mathrm{E}\{\frac{\partial^2 l_c(\theta)}{\partial\theta^2} \mid \mathrm{y}, \mathrm{w}\} - \mathrm{Var}\{\frac{\partial l_c(\theta)}{\partial\theta} \mid \mathrm{y}, \mathrm{w}\}.$$

Then, by applying the quadrature approximations of the E-step,

$$\mathrm{E}\{\frac{\partial^2 l_c(\hat{\theta})}{\partial\theta^2} \mid \mathrm{y}, \mathrm{w}\} \approx \sum_{i=1}^{n}\sum_{j=1}^{M} \mathrm{A}_{ij}^{(\mathrm{T})} \frac{\partial^2 l_{ci}(\theta^{(T)})}{\partial\theta^2} = \frac{\partial^2 \mathrm{Q}(\theta|\theta^{(T)})}{\partial\theta^2}\Big|_{\theta=\theta^{(\mathrm{T})}} \qquad (6)$$

and

$$\mathrm{Var}\{\frac{\partial l_c(\hat{\theta})}{\partial\theta} \mid \mathrm{y}, \mathrm{w}\} = \mathrm{E}\{\left(\frac{\partial l_c(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l_c(\hat{\theta})}{\partial\theta}\right)' \mid \mathrm{y}, \mathrm{w}\},$$

since

$$\mathrm{E}\{\left(\frac{\partial l_c(\hat{\theta})}{\partial\theta}\right) \mid \mathrm{y}, \mathrm{w}\} = 0.$$

Therefore,

$$\mathrm{E}\{\left(\frac{\partial l_c(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l_c(\hat{\theta})}{\partial\theta}\right)' \mid \mathrm{y}, \mathrm{w}\} = \sum_{i=1}^{n}\mathrm{E}\{\left(\frac{\partial l_{ci}(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial l_{ci}(\hat{\theta})}{\partial\theta}\right)' \mid \mathrm{y}, \mathrm{w}\}$$

$$+ \sum_{i\neq j}\mathrm{E}\{\left(\frac{\partial l_{ci}(\hat{\theta})}{\partial\theta}\right) \mid \mathrm{y}, \mathrm{w}\}\,\mathrm{E}\{\left(\frac{\partial l_{cj}(\hat{\theta})}{\partial\theta}\right)' \mid \mathrm{y}, \mathrm{w}\}$$

since for $i \neq j$ the cases are independent. Also,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\{\left(\frac{\partial l_{ci}(\hat{\theta})}{\partial\theta}\right) \mid \mathrm{y}, \mathrm{w}\}\,\mathrm{E}\{\left(\frac{\partial l_{cj}(\hat{\theta})}{\partial\theta}\right)' \mid \mathrm{y}, \mathrm{w}\} = 0,$$

therefore,

$$\sum_{i\neq j} \mathrm{E}\{\left(\frac{\partial l_{ci}(\widehat{\theta})}{\partial \theta}\right) | \mathrm{y}, \mathrm{w}\}\, \mathrm{E}\{\left(\frac{\partial l_{cj}(\widehat{\theta})}{\partial \theta}\right)' | \mathrm{y}, \mathrm{w}\}$$

$$= -\sum_{i=1}^{n} \mathrm{E}\{\left(\frac{\partial l_{ci}(\widehat{\theta})}{\partial \theta}\right) | \mathrm{y}, \mathrm{w}\}\, \mathrm{E}\{\left(\frac{\partial l_{ci}(\widehat{\theta})}{\partial \theta}\right)' | \mathrm{y}, \mathrm{w}\}.$$

This leads to the approximation

$$\mathrm{E}\{\left(\frac{\partial l_c(\widehat{\theta})}{\partial \theta}\right)\left(\frac{\partial l_c(\widehat{\theta})}{\partial \theta}\right)' | \mathrm{y}, \mathrm{w}\} \approx \sum_{i=1}^{n}\sum_{j=1}^{M} \mathrm{A}_{ij}^{(\mathrm{T})} \left(\frac{\partial l_{ci}(\theta^{(T)})}{\partial \theta}\right)\left(\frac{\partial l_{ci}(\theta^{(T)})}{\partial \theta}\right)'$$

$$- \sum_{i=1}^{n}\left(\sum_{k=1}^{M} \mathrm{A}_{ik}^{(\mathrm{T})} \frac{\partial l_{ci}(\theta^{(T)})}{\partial \theta}\right)\left(\sum_{k=1}^{M} \mathrm{A}_{ik}^{(\mathrm{T})} \frac{\partial l_{ci}(\theta^{(T)})}{\partial \theta}\right)' \qquad (7)$$

where, the terms are as defined previously. Although, only the covariance submatrix for θ_1 may be of interest, the entire matrix must be calculated: while (6) may be block diagonal with respect to θ_1, θ_2, and θ_3 (7) is not, thus the observed information submatrix for θ_1 cannot be inverted separately. This means the calculation could be quite cumbersome, especially if the dimension of θ is large. The first term (6) of the expression for the information matrix should be relatively easy to calculate since its identical to the formulation of observed information matrix for the weighted log-likelihood maximized in the M-step. The second term (7) will be more difficult to calculate.

The previous results were again in terms of situations 1 and 3 but modifications for other types of extra information are quite simple. The expressions for (6) and (7) for external replication data (situation 5) are basically the same, the only difference being that the first n_1 cases (the external data) will not involve y. Therefore, summation involving derivatives with respect to θ_1 will be from n_1+ 1 to n. When there is validation data (situations 2 and 4) one only need add minus the observed information for the n_1

validation cases, $\sum_{i=1}^{n_1} \frac{\partial^2 l_{ci}(\theta^{(T)})}{\partial \theta^2}$, to the expressions for (6) and (7). Except, in this case the expressions for (6) and (7) sum from $n_1 + 1$ to n.

3.5.5 Finding Initial Estimates

Finding reasonable initial estimates can usually be achieved through the use of naive or moment based estimators. To find initial estimates for the parameters in the distribution of y (1), θ_1, in many cases it may be reasonable to simply use the naive estimators which would be used if one assumed that there were no measurement error. However, if better estimates are required to achieve a faster rate of convergence then one might use a moment based approach such as the "regression calibration" method (See Carroll et. al., 1995). To find an initial estimate for the parameter vector in the measurement error distribution (2), θ_2, one must take advantage of the extra information. For example, if the extra information is in the form of replication (situations 3 and 5) and w|x ~ N(x, σ_w^2), then a reasonable initial estimate would be

$$\widehat{\sigma}_w^{2(0)} = \sum_{i=1}^{n}\sum_{j=1}^{r_i}(w_{ij} - \overline{w}_i)^2/(\sum_{i=1}^{n} r_i - n)$$

where, $\overline{w}_i = \sum_{j=1}^{r_i} w_{ij}/r_i$. When the extra information is validation data (situations 2 and 4) the estimate need only be modified slightly,

$$\widehat{\sigma}_w^{2(0)} = \sum_{i=1}^{n_1}(w_i - x_i)^2/n_1$$

where, n_1 is the number of cases with validation data.

Finding initial estimates for the parameters in the distribution of x (3), θ_3, may be slightly more difficult, since most often in measurement error problems the actual values of x are missing. So, one must take advantage of model assumptions, for example if one

assumes an additive measurement error model where the measurement w = x + u, where x and u are independent and u has mean 0. Then $\mu_x = \mu_w$, therefore a reasonable initial estimate is $\widehat{\mu}_x^{(0)} = \sum_{i=1}^{n}\sum_{j=1}^{r_i} w_{ij} / \sum_{i=1}^{n} r_i$. Under these model assumptions Var(x) = Var(w) − Var(u), and this leads to

$$\widehat{\sigma}_x^{2(0)} = \{\sum_{i=1}^{n}\sum_{j=1}^{r_i}(w_{ij} - \overline{w})^2 - \sum_{i=1}^{n}\sum_{j=1}^{r_i}(w_{ij} - \overline{w}_i)^2\}/(2n - \sum_{i=1}^{n} r_i)$$

where, $\overline{w} = \sum_{i=1}^{n}\sum_{j=1}^{r_i} w_{ij} / \sum_{i=1}^{n} r_i$. When x does not have a normal distribution these estimates can be used to obtain method of moment estimators for parameters. For example if x has a gamma distribution an initial estimate of the shape parameter is $\widehat{\alpha}^{(0)} = \{\widehat{\mu}^{(0)}\}^2 / \widehat{\sigma}_x^{2(0)}$ and the scale parameter is $\widehat{\beta}^{(0)} = \widehat{\sigma}_x^{2(0)} / \widehat{\mu}^{(0)}$. When validation data is available this can be used in the usual manner to find initial estimate for the parameters in the distribution of x.

3.5.6 Extensions

The model presented thus far only involves a single explanatory variable measured with error and the possible inclusion of explanatory variables measured without error. It is fairly straightforward to include a quadratic term for the explanatory variable measured with error or interaction terms with the explanatory variables measured without error. These terms only involve the response distribution (1) and will create only slight increases in computational difficulty in the M-step and for calculating the transformation for the Gauss-Hermite quadrature. Expanding the model to include more than one explanatory variable measured with error is considerably more difficult. Multidimensional quadrature is quite cumbersome so an approach like Kuha (1996), where Monte-Carlo integration is used in the E-step may be more reasonable, and this approach also results in similar likelihood calculations in the M-step. However, for integrals in one dimension, Monte-Carlo integration may require a quite large Monte-

Carlo sample size in order to achieve the same degree of accuracy as several nodes of Gauss-Hermite quadrature.

3.6 Example

The data used here come from a study by Clayton (1991), in which the ratio of polyunsaturated to saturated fat intake (P/S) was related to death by heart disease. This ratio was measured on 336 male subjects by a one week dietary survey. The survey was repeated for a subset of 76 subjects six months later. The response variable was 1 if the subject died from heart disease during the study and 0 otherwise. Additionally, the age of the subject at the beginning of study was also accounted for. This study can be considered a regression problem with explanatory variable measurement error since the true value P/S might be thought of as a long term average of the ratio of polyunsaturated to saturated fat in the diet. A measurement of it is taken from the one week survey, but it contains error since a subject's diet will vary from week to week. Extra information is in the form of internal replication (situation 3) if it is assumed that the second measurements of P/S six months later are true replicates. The model used here to analyze the data will assume the response has a binary logistic regression on log P/S and age. Examination of a normal probability plot of residuals from the regression of the log of the P/S measurements on age in Figure 3.1 shows that it is consistent with a model where the measurement error, w | x is taken to have a $N(x, \sigma_w^2)$ distribution, and x, the true value of log P/S is taken to have a $N(\gamma_0 + \gamma_1 z, \sigma_x^2)$ distribution, where z is the age of the subject. Replicate measurements allow for a check of the assumption of normality for the measurement error distribution by a normal probability plot of $\{(\frac{1}{\sqrt{2}}(w_{i2} - w_{i1}), \ldots, \sqrt{\frac{r_i}{r_i+1}}(w_{ir_i} - \frac{1}{r_i-1}\sum_{j=1}^{r_i-1} w_{ij})\}$ for the cases with replicate measurements. The plot for the measurements of log P/S in Figure 3.2 indicates the possibility of heavy tails in the distribution of the measurement error. Higdon and Schafer (1998) show that this situation can present difficulties especially if the measurement error is large. For purposes of

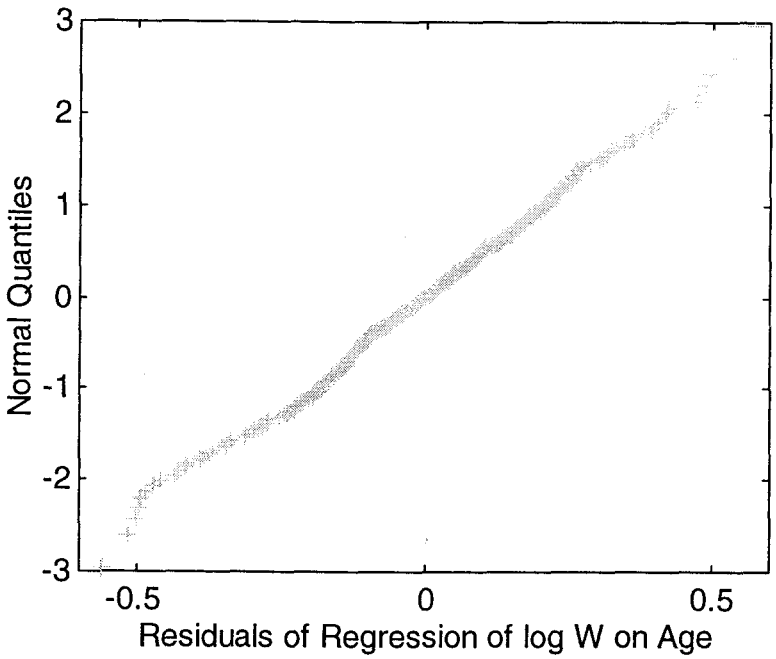

Figure 3.1. Normal probability plot of the residuals of the regression of the log of the measurements of P/S on age.

demonstrating these techniques, however, we will proceed as if the assumption of normality for the measurement error is reasonable.

The initial estimates of the parameters in the logistic regression on log (P/S) and age based ignoring the measurement error are

$$\text{logit}(p) = 6.59 - 2.55 \times \log(P/S) + .0035 \times \text{age}.$$

The initial estimate for measurement error variance σ_w^2 is 0.012, and the initial estimates for E(x | z) and σ_x^2 are $4.30 - .0002 \times$ age and 0.027 respectively. The estimates of the regression coefficient of log(P/S) for each iteration of the EM algorithm is given Table 3.2.

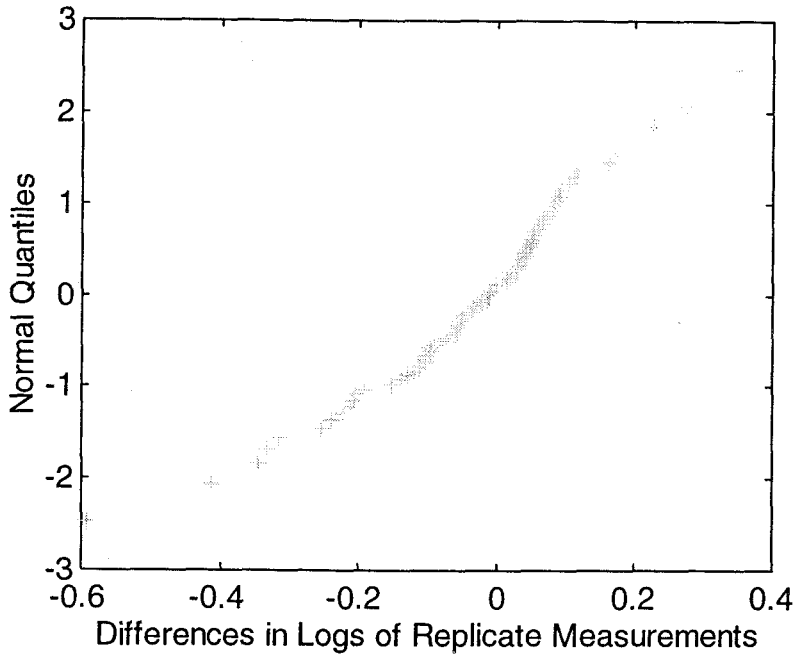

Figure 3.2. Normal probability plot of the differences of the logs of replicate measurements of P/S.

Table 3.2. Estimates at Iterations of the EM Algorithm.

Iteration	*Estimate of Coefficient of log(P/S)*
0	*-2.5497*
1	*-3.2756*
2	*-3.5228*
3	*-3.6136*
5	*-3.6718*
10	*-3.7123*
15	*-3.7288*

The algorithm converged after 15 iterations with a relative rate of convergence of .001.

It is estimated that measurement error accounts for 33% of the variation in the measurements of log(P/S) (i.e. $\widehat{\sigma}_w^2/(\widehat{\sigma}_w^2 + \widehat{\sigma}_x^2) = 0.33$). Results of this analysis show that the fitted model for the log-odds for death from heart disease is

$$\text{logit}(p) = 11.63 - 3.73 \times \log(P/S) + .0032 \times \text{age}.$$

There is a significant association between P/S and the odds of dying from heart disease after adjusting for age, based on the p-value of 0.003 from the likelihood ratio test. A 95% confidence interval for the coefficient of log(P/S) was estimated to be -1.25 to -6.75. This is found by inverting the likelihood ratio test using a trial and error method of fitting reduced models for various fixed values of the regression coefficient and finding the values at which the p-value for the test was just less than .05. A starting point for the trial and error method was the confidence interval of -1.08 to -6.48 based on the assumption of normality and the standard error estimate from Section 3.5.6 of 1.35. There was no evidence of an interaction between log(P/S) and age (p-value = 0.97) and there was suggestive evidence that the relationship between log(P/S) and the log-odds of death from heart disease was not linear (p-value = 0.062 for the quadratic term).

The analysis was implemented by using a program generated in Matlab. The fitting of various models using a relative rate of convergence of .001 took between 15 and 47 iterations and between 10 and 65 seconds to converge on a Sun SPARC 20 workstation.

3.7 References

Abramowitz, M., and Stegun J. (1972), *Handbook of Mathematical Functions*, New York: Dover.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, New York: Chapman and Hall.

Clayton, D. G. (1991), "Models for the Analysis of Cohort and Case-control Studies with Inaccurately Measured Data," In Dwyer, J. H., Feinleib, M., Lippert, P., and] Hoffmeister, H. (Editors), *Statistical Models for Longitudinal Studies*, New York: Oxford University Press, 301-331.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

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

Higdon R., and Schafer, D. W. (1998), "Likelihood Analysis for Regression with Measurement Error," Oregon State University, Unpublished Manuscript.

Kuchenoff, H., and Carroll, R. J. (1997), "Biases in Segmented Regression with Errors in Predictors," *Statistics in Medicine*, 439-450.

Kuha, J. (1996), "Monte Carlo Methods for Adjusting for Measurement Error," *The XVIIth International Biometric Conference.*

Liu, Q., and Pierce, D. A. (1994), "A Note on Gauss-Hermite Quadrature," *Biometrika*, 81, 624-629

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

Schafer, D. W. (1987), "Covariate Measurement Error in Generalized Linear Models," *Biometrika*, 74, 385-391.

Schafer, D. W. (1993), "Likelihood Analysis for Probit Regression with Measurement Errors," *Biometrika*, 80, 899-904.

Schafer, D. W., and Purdy, K. G. (1996), "Likelihood Analysis for Errors-in-Variables Regression with Replicate Measurements," *Biometrika*, 83, 813-824.

4. CONCLUSION

This main goal of this thesis is to present likelihood analysis as a reasonable tool for regression problems with explanatory variable measurement error. This has been done by presenting (i) a general structural likelihood model, (ii) the numerical tools by which maximum likelihood estimators and likelihood ratio statistics may be found, and (iii) a simulation study which demonstrates some of the efficiency, validity of confidence intervals, and robustness characteristics of maximum likelihood estimators relative to other popular methods.

4.1 Computation

The structural model presented in this thesis allows for likelihood analysis under the assumptions that the individual cases are independent, there is "non-differential" measurement error, and that there is some form of extra information about the measurement error distribution. The first paper (Chapter 2) only considers the case where extra information is in the form of replicate measurement, but the second paper (Chapter 3) generalizes the approach for other types of extra information. The numerical approach of using the EM algorithm with modified Gauss-Hermite quadrature to approximate integrals in the E-step reasonably allows one to do likelihood analysis when a single explanatory variable is measured with error. Although one could attempt to maximize the observed data log-likelihood directly there are some nice computational advantages to using the EM algorithm. First, the use of Gauss-Hermite quadrature in the E-step results in relatively straightforward ways to maximize weighted log-likelihoods in the M-step. Second, the maximized value of the log-likelihood for use in likelihood ratio inference is a by-product of the algorithm, and third, Louis's method allows for the calculation of asymptotic standard errors. The computation for the standard errors can be a bit cumbersome, and in light of the advantages of likelihood ratio inference of asymptotic

normality plus standard errors presented in this thesis the value of this is suspect. None the less, the option is available.

This thesis does not propose that that computation methods presented here are necessarily the only or the best to maximize the observed data log-likelihood, only that the computational methods presented here are reasonable and work well in practice. Certainly, a reasonable alternative would be to maximize the observed data log-likelihood directly using a non-linear optimization routine. One of the principal advantages in this approach would be in using pre-existing programs for the maximization. However, this approach would still require E-step like calculations to evaluate the observed data log-likelihood, and since the M-step is a relatively straightforward programming problem, the advantages would not be great. Another advantage might be in speed, since the EM algorithm is often quite slow too converge. It should be noted however, that a couple of pre-existing optimization programs were a good deal slower than the EM algorithm when applied to some of the problems presented in this thesis.

4.2 Simulations

The examinations of robustness, efficiency, and validity are necessarily limited in scope since only linear and logistic regression with extra information in the form of internal replication were examined along with mostly one distribution at a time departures from the "everything normal" model. There were however a number of consistent patterns in the relative mean square errors and coverage error rates of 95% confidence intervals, over different amounts of replication and different measurement error sizes, suggesting the results might extend to other situations as well.

The maximum likelihood estimator based on the "everything normal" model does appear to be more robust than previously suspected. It appears to be satisfactory when the distribution of x is skewed or heavy tailed, but not so good when the distribution of the measurement error is heavy tailed or multiplicative in nature. Surprisingly, the maximum likelihood estimator based on the "everything normal" model appears to be as or more

robust to these departures in normality than moment based estimators using weaker assumptions. Likelihood analysis can offer more efficiency and flexibility by modeling the components to the structural measurement error model, for example, by specifying the distribution of x to be gamma, lognormal, or a mixture of normals. However, the simulations do suggest that when the measurement error distribution has heavy tails the subsequent analysis could be risky with any method.

The simulations also demonstrate the poor performance of confidence intervals based on approximate normality and asymptotic standard errors. Additionally, the performance of moment based method were generally no better and often much worse than the maximum likelihood estimator based on the "everything normal" model.

4.3 Extensions to the Structural Model

The addition of quadratic, x^2, terms or product terms, $x * z$, is not difficult with the structural likelihood model presented in this thesis. In moment methods, by contrast, it would be necessary to specify the marginal distribution and the measurement error distribution (or at least first and second moments of these) separately for x and for the terms constructed from x. Extending the models presented in this thesis to more than one explanatory variable measured with error is easy in principal. However, the E-step would then require approximation of multiple integrals and this would be come a considerable numerical hindrance using Gauss-Hermite quadrature. In this case Kuha's (1996) approach of using Monte Carlo integration might be more attractive. Markov Chain Monte Carlo methods might also be useful in this situation if the appropriate conditional distributions could be specified.

It is clear that further investigation into the use of likelihood methods for regression models with explanatory variable measurement error is warranted. It is certainly of interest to know if the results presented here hold for other types of extra information and other types of models, and if these methods can be extended to more complicated models (multiple explanatory variables measured with error, correlated data,

random effects, etc.). Better methods for checking model assumptions need to be developed if likelihood methods are to gain widespread acceptance for analysis of measurement error models. In any event this thesis has made the point that likelihood should at least be considered in regression problems with measurement error, especially if the underlying distributions can be modeled adequately. There are some definite advantages to likelihood analysis which cannot be ignored, especially the advantage of the likelihood ratio statistic over approximate normality assumptions plus standard errors for inference and confidence intervals.

BIBLIOGRAPHY

Abramowitz, M., and Stegun J. (1972), *Handbook of Mathematical functions*, New York: Dover.

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

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

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, New York: Chapman and Hall.

Carroll, R. J., Spiegelman, C., Lan, K. K., and Abbott, R. D. (1984), "On Errors-in-Variables for Binary Regression Models," *Biometrika*, 71, 19-26.

Clayton, D. G. (1991), "Models for the Analysis of Cohort and Case-control Studies with Inaccurately Measured Data," In Dwyer, J. H., Feinleib, M., Lippert, P., and Hoffmeister, H. (Editors), *Statistical Models for Longitudinal Studies*, New York: Oxford University Press, 301-331.

Crouch E. A. and Spiegelman D. (1990), "The Evaluation of Integrals of the Form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)dt$: Applications to Logistic-Normal Models," *Journal of the American Statistical Association*, 85, 464-467.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion)," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

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

Higdon R., and Schafer, D. W. (1998a), "Likelihood Analysis for Regression with Measurement Error," Oregon State University, Unpublished Manuscript.

Higdon R., and Schafer, D. W. (1998b), "Maximum Likelihood Computations for Regression with Measurement Error," Oregon State University, Unpublished Manuscript.

Kendall, M. G., and Stuart, A. (1979) *The Advanced Theory of Statistics, Vol. 2*, London: Griffin.

Kuchenoff, H., and Carroll, R. J. (1997), "Biases in Segmented Regression with Errors in Predictors," *Statistics in Medicine*, 439-450.

Kuha, J. (1996), "Monte Carlo Methods for Adjusting for Measurement Error," *The XVIIth International Biometric Conference*.

Liu, Q., and Pierce, D. A. (1994), "A Note on Gauss-Hermite Quadrature," *Biometrika*, 81, 624-629.

Liu, X., and Liang, K. Y. (1992), "Efficacy of Repeated Measures in Regression Models with Measurement Error," *Biometrics*, 48, 645-654.

Mandansky, A. (1959), "The Fitting of Straight Lines when Both Variables are Subject to Error," *Journal of the American Statistical Association*, 54, 173-205.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 29-265.

Schafer, D. W. (1987), "Covariate Measurement Error in Generalized Linear Models," *Biometrika*, 74, 385-391.

Schafer, D. W. (1992), "Replacement Methods for Measurement Error Models," *Oregon State University Technical Report 151*.

Schafer, D. W. (1993), "Likelihood Analysis for Probit Regression with Measurement Errors," *Biometrika*, 80, 899-904.

Schafer, D. W., and Purdy, K. G. (1996), "Likelihood Analysis for Errors-in-Variables Regression with Replicate Measurements," *Biometrika*, 83, 813-824.

Stefanski, L. A., and Carroll, R. J. (1985) "Covariate Measurement Error in Logistic Regression," *Annals of Statistics*, 8, 407-422.

Zhao, Y., and Lee, A. H. (1997), "A Simulation Study of Estimators for Generalized Linear Measurement Error Models," Unpublished Manuscript.

APPENDIX

This appendix presents tables of simulation results for the 9 different simulations and corresponding figures in presented in Chapter 2. The details of the simulation conditions are presented in the appendix of Chapter 2 (Section 2.8). The tables report the observed bias, mean square error (MSE), proportion of "left-side misses" of a 95% percent confidence interval (P(UL< β)) and the proportion of "right-side misses" of 95% confidence interval (P(LL > β)), for each of the estimators over the Monte-Carlo distribution. The estimators are as follows: $\widehat{\beta}_{norm}$; the maximum likelihood estimator assuming x and the measurement error have normal distributions, $\widehat{\beta}_{gam}$; the maximum likelihood estimator assuming x has a gamma distribution and the measurement error has a normal distribution, $\widehat{\beta}_{naive}$; the usual estimator that would be used when there is no measurement error, $\widehat{\beta}_{f}$; the estimator based on Fuller's method-of-moments, $\widehat{\beta}_{nmix}$; the maximum likelihood estimator assuming x is distrubuted as a mixture of normals and the measurement error has a normal distribution, $\widehat{\beta}_{mixn}$; the maximum likelihood estimator assuming x has a normal distribution and the measurement error is distributed as a mixture of normals, $\widehat{\beta}_{cal}$; the "regression calibration" estimator, and $\widehat{\beta}_{mcal}$; a multiplicative "regression calibration" estimator. Also, $\widehat{\beta}_{cal\text{-}se}$ and $\widehat{\beta}_{cal\text{-}lr}$ represent the "regression calibration" estimator using an asymptotic standard error and a pseudo-likelihood ratio statistic respectively to compute confidence intervals. Otherwise, $\widehat{\beta}_{cal}$ uses asymtotic standard errors to compute confidence intervals.

Table A.1. Simulation Results Normal-Normal-Gamma Model

			Replication 20%				Replication 5%			
			n = 100							
a,b	σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
4,8	60	$\hat{\beta}_{gam}$	.0012	.0051	.029	.034	.0045	.0070	.035	.034
		$\hat{\beta}_{norm}$	-.0002	.0053	.035	.034	-.0071	.0078	.034	.018
		$\hat{\beta}_{naive}$	-.0968	.0122	.501	0	-.0968	.0119	.504	0
		$\hat{\beta}_f$	.0048	.0075	.045	.017	.0030	.0106	.092	0
	240	$\hat{\beta}_{gam}$	-.0007	.0152	.026	.034	-.0035	.0192	.020	.029
		$\hat{\beta}_{norm}$	-.0144	.0177	.017	.021	-.0738	.0224	.039	.001
		$\hat{\beta}_{naive}$	-.2446	.0619	.999	0	-.2437	.0615	1	0
		$\hat{\beta}_f$	.0130	.0299	.128	0	-.0944	.0472	.270	.002

a,b	σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
1,18	60	$\hat{\beta}_{gam}$	.0076	.0031	.014	.027	.0019	.0033	.022	.023
		$\hat{\beta}_{norm}$	.0027	.0034	.019	.027	-.0087	.0060	.032	.018
		$\hat{\beta}_{naive}$	-.0787	.0085	.417	0	-.0846	.0096	.469	0
		$\hat{\beta}_f$	.0069	.0051	.032	.013	-.002	.0079	.079	.002
	240	$\hat{\beta}_{gam}$	.0217	.0091	.016	.046	.0248	.0112	.019	.040
		$\hat{\beta}_{norm}$	-.0079	.0127	.025	.012	-.0595	.0184	.044	.003
		$\hat{\beta}_{naive}$	-.2167	.0496	.997	0	-.2186	.0503	.997	0
		$\hat{\beta}_f$	.0188	.0232	.093	0	-.0724	.0323	.262	.002

Table A.1. (Continued)

			n = 500							
a,b	σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
4,8	60	$\widehat{\beta}_{gam}$	-.0036	.0008	.034	.025	-.0033	.0010	.020	.027
		$\widehat{\beta}_{norm}$	-.0045	.0009	.030	.019	-.0107	.0014	.026	.023
		$\widehat{\beta}_{naive}$	-.0883	.0083	.988	0	-.0936	.0093	.990	0
		$\widehat{\beta}_{f}$	-.0017	.0009	.048	.020	-.0003	.0018	.080	.012
	240	$\widehat{\beta}_{gam}$	-.0166	.0024	.019	.023	-.025	.0032	.023	.018
		$\widehat{\beta}_{norm}$	-.0302	.0038	.024	.014	-.0749	.0087	.026	.002
		$\widehat{\beta}_{naive}$	-.2297	.0532	1	0	-.239	.0576	1	0
		$\widehat{\beta}_{f}$	-.0065	.0046	.117	.002	-.0151	.0130	.173	0

a,b	σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
1,18	60	$\widehat{\beta}_{gam}$	.0014	0006	.027	.033	.0023	.0006	.025	.026
		$\widehat{\beta}_{norm}$	-.0032	.0007	.032	.034	-.0078	.0009	.032	.019
		$\widehat{\beta}_{naive}$	-.0727	.0057	.938	0	-.0774	.0064	.976	0
		$\widehat{\beta}_{f}$	-.0011	.0008	.047	.032	-.0002	.0012	.062	.022
	240	$\widehat{\beta}_{gam}$	.0112	.0013	.013	.046	.0114	.0016	.014	.046
		$\widehat{\beta}_{norm}$	-.0201	.0026	.023	.022	-.0587	.0061	.021	0
		$\widehat{\beta}_{naive}$	-.2007	.0407	1	0	-.2109	.0450	1	0
		$\widehat{\beta}_{f}$	-.001	.0032	.095	.111	-.0107	.0087	.157	.004

Table A.2. Simulation Results Normal-Normal-t Model

		Replication 20%				Replication 5%			
				n = 100					
σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
60	$\hat{\beta}_{nmix}$	.0103	.0080	.028	.034	.0181	.0160	.024	.044
	$\hat{\beta}_{norm}$	.0113	.0090	.026	.025	.0221	.0182	.022	.020
	$\hat{\beta}_{naive}$	-.1030	.0140	.479	0	-.109	.0155	.542	0
	$\hat{\beta}_f$	.0009	.0082	.060	.024	-.0021	.0184	.094	.021
240	$\hat{\beta}_{nmix}$	.0393	.0432	.020	.022	.0136	.0435	.045.	.013
	$\hat{\beta}_{norm}$	.0364	.0399	.023	.023	-.0212	.0397	.043	.003
	$\hat{\beta}_{naive}$	-.2545	.0691	.994	0	-.2655	.0747	.995	0
	$\hat{\beta}_f$	-.0518	.0325	.167	.002	-.1217	.0613	.292	.005
				n = 500					
σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
60	$\hat{\beta}_{nmix}$	.0019	.0010	.022	.034	.0007	.0016	.029	.023
	$\hat{\beta}_{norm}$	.0017	.0012	.024	.032	.0012	.0026	.029	.028
	$\hat{\beta}_{naive}$	-.0969	.1020	.978	0	-.1052	.0120	.985	0
	$\hat{\beta}_f$	.0006	.0011	.042	.032	.001	.0025	.079	.018
240	$\hat{\beta}_{nmix}$	.0041	.0032	.020	.031	.0032	.0056	.026	.033
	$\hat{\beta}_{norm}$	.0022	.006	.026	.025	-.0135	.0123	.039	.010
	$\hat{\beta}_{naive}$	-.2440	.0608	.997	0	-.2544	.0659	1	0
	$\hat{\beta}_f$	-.0068	.0054	.112	.009	-.0225	.0166	.173	.001

Table A.3. Simulation Results Normal-t-Normal Model

		Replication 20%				Replication 5%			
		n = 100							
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
60	$\hat{\beta}_{mixn}$	.0073	.0100	.033	.031	.0190	.0374	.081	.038
	$\hat{\beta}_{nmix}$	.0003	.0091	.072	.046	-.0054	.0171	.093	.035
	$\hat{\beta}_{norm}$	.0014	.0083	.061	.043	-.0056	.0145	.068	.034
	$\hat{\beta}_{naive}$	-.0872	.0110	.470	0	-.0910	.0121	.438	.001
	$\hat{\beta}_f$	-.0113	.0082	.095	.018	-.0328	.0162	.152	.004
240	$\hat{\beta}_{mixn}$	.1013	.0544	.040	.053	.1691	.1301	.032.	.036
	$\hat{\beta}_{nmix}$	-.0080	.0442	.180	.050	-.0655	.0537	.287	.019
	$\hat{\beta}_{norm}$	.0123	.0329	.080	.047	-.0446	.0402	.100	.014
	$\hat{\beta}_{naive}$	-.2204	.0529	.976	0	-.2349	.0592	.990	0
	$\hat{\beta}_f$	-.0728	.0316	.271	.001	-.1391	.0525	.402	0
		n = 500							
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
60	$\hat{\beta}_{mixn}$	.0074	.0011	.025	.018	.0135	.0126	.074	.013
	$\hat{\beta}_{nmix}$	-.0052	.0031	.121	.060	-.0101	.0056	.177	.064
	$\hat{\beta}_{norm}$	-.0008	.0027	.091	.065	-.0032	.0053	.108	.074
	$\hat{\beta}_{naive}$	-.0892	.0090	.961	0	-.0949	.0099	.986	0
	$\hat{\beta}_f$	-.0064	.0028	.133	.036	-.0109	.0050	.178	.021
240	$\hat{\beta}_{mixn}$	.0540	.0083	.015	.018	.2753	.1013	.016	.024
	$\hat{\beta}_{nmix}$	-.0571	.0191	.362	.043	-.1064	.0326	.532	.037
	$\hat{\beta}_{norm}$	-.0096	.0106	.125	.062	-.0436	.0194	.132	.049
	$\hat{\beta}_{naive}$	-.2297	.0541	1	0	-.2394	.0587	1	0
	$\hat{\beta}_f$	-.0443	.0125	.271	.002	-.0808	.0259	.402	.002

Table A.4. Simulation Results Normal-Logormal-Lognormal Model

X ~ Lognormal(4,.05)

n = 100

		Replication 20%				Replication 5%			
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	.0081	.0069	.020	.030	.0191	.0169	.022	.033
	$\hat{\beta}_{norm}$	.0060	.0073	.026	.028	.0104	.0151	.027	.037
	$\hat{\beta}_{naive}$	-.0790	.0099	.271	0	-.0862	.0114	.312	0
	$\hat{\beta}_{mcal}$	.0084	.0067	.044	.059	.0511	.0292	.068	.132
	$\hat{\beta}_f$	.0010	.0071	.036	.024	-.0040	.0141	.084	.011
.08	$\hat{\beta}_{lognor}$	.0443	.0575	.030	.033	-.0127	.0615	.035	.012
	$\hat{\beta}_{norm}$	.0265	.0558	.058	.013	-.0618	.0581	.052	.001
	$\hat{\beta}_{naive}$	-.3090	.0977	.999	0	-.3185	.1036	1	0
	$\hat{\beta}_{mcal}$	-.1276	.0788	.326	.030	-.3081	.2649	.520	.026
	$\hat{\beta}_f$	-.1125	.0515	.270	0	-.1868	.1056	.442	0

n = 500

σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	-.0001	.0013	.033	.030	-.0009	.0018	.027	.018
	$\hat{\beta}_{norm}$	-.0026	.0014	.043	.029	-.0040	.0020	.041	.020
	$\hat{\beta}_{naive}$	-.0801	.0072	.854	0	-.0867	.0082	.894	0
	$\hat{\beta}_{mcal}$	.0040	.0014	.030	.057	.0047	.0020	.057	.096
	$\hat{\beta}_f$	-.0034	.0014	.056	.025	-.0037	.0021	.082	.014
.08	$\hat{\beta}_{lognor}$	-.0004	.0095	.021	.022	-.0165	.0177	.025	.019
	$\hat{\beta}_{norm}$	-.0178	.0125	.053	.019	-.0506	.0212	.047	.005
	$\hat{\beta}_{naive}$	-.3087	.0957	1	0	-.3207	.1033	1	0
	$\hat{\beta}_{mcal}$	-.0030	.0107	.209	.104	-.0185	.0958	.368	.199
	$\hat{\beta}_f$	-.0476	.0163	.233	0	-.0945	.0416	.318	0

Table A.4. (Continued)

X ~ Lognormal(4,.5)

n = 100

		Replication 20%				Replication 5%			
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	.0005	.0005	.019	.027	-.0008	.0005	.038	.029
	$\hat{\beta}_{norm}$	-.0020	.0007	.133	.086	-.0040	.0008	.130	.076
	$\hat{\beta}_{naive}$	-.0135	.0008	.227	.031	-.0146	.0008	.250	.032
	$\hat{\beta}_{mcal}$	.0027	.0006	.058	.065	.0074	.0009	.071	.099
	$\hat{\beta}_f$	-.0030	.0008	.169	.108	-.0056	.0008	.175	.066
.08	$\hat{\beta}_{lognor}$	.0012	.0015	.023	.027	-.0005	.0018	.028	.029
	$\hat{\beta}_{norm}$	-.0251	.0048	.278	.077	-.0455	.0073	.292	.050
	$\hat{\beta}_{naive}$	-.0919	.0107	.826	.003	-.0968	.0120	.851	.005
	$\hat{\beta}_{mcal}$	.0129	.0046	.125	.097	.0311	.0169	.148	.144
	$\hat{\beta}_f$	-.0385	.0061	.404	.043	-.0541	.0090	.481	.022

n = 500

σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	.0004	.00009	.022	.028	-.0001	.00009	.025	.025
	$\hat{\beta}_{norm}$	-.0019	.00016	.154	.088	-.0029	.00022	.199	.097
	$\hat{\beta}_{naive}$	-.0130	.00031	.475	.008	-.0143	.00036	.522	.012
	$\hat{\beta}_{mcal}$	.0005	.00015	.050	.030	.0015	.00016	.040	.048
	$\hat{\beta}_f$	-.0024	.00017	.178	.092	-.0031	.00024	.229	.105
.08	$\hat{\beta}_{lognor}$	-.0020	.0003	.023	.021	-.0030	.0004	.029	.028
	$\hat{\beta}_{norm}$	-.0245	.0019	.377	.049	-.0377	.0030	.364	.033
	$\hat{\beta}_{naive}$	-.0954	.0098	.997	0	-.1018	.0110	.999	0
	$\hat{\beta}_{mcal}$	.0016	.0011	.080	.059	.0058	.0022	.152	.136
	$\hat{\beta}_f$	-.0315	.0026	.536	.036	-.0436	.0041	.561	.011

Table A.5. Simulation Results Logistic Normal-Normal Model

		Replication 20%				Replication 5%			
				n = 300					
σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{norm}$	-.0268	2.133	.019	.023	-.0356	2.486	.027	.024
	$\hat{\beta}_{cal\text{-}se}$	-.0086	2.051	.021	.030	-.0494	2.475	.020	.037
	$\hat{\beta}_{cal\text{-}lr}$			.048	.025			.047	.037
	$\hat{\beta}_{naive}$	.7365	1.896	.008	.099	.7663	2.030	.006	.100
.02	$\hat{\beta}_{norm}$	-.0211	4.955	.040	.034	.0407	5.720	.020	.029
	$\hat{\beta}_{cal\text{-}se}$	-.0191	4.380	.005	.052	-1.862	1064	0	.067
	$\hat{\beta}_{cal\text{-}lr}$			.047	.036			.066	.108
	$\hat{\beta}_{naive}$	1.940	4.723	0	.568	1.950	4.576	0	.577
				n = 1000					
σ_w^2	Estimator	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{norm}$	-.0299	.6298	.028	.032	-.0182	..6368	.015	.020
	$\hat{\beta}_{cal\text{-}se}$	-.0202	.6086	.026	.030	-.0280	.6261	.022	.023
	$\hat{\beta}_{cal\text{-}lr}$			.037	.041			.048	.035
	$\hat{\beta}_{naive}$	.7198	.9203	.002	.223	.7669	.9703	0	.239
.02	$\hat{\beta}_{norm}$	.1101	1.087	.036	.026	.4134	1.417	.024	.036
	$\hat{\beta}_{cal\text{-}se}$	.0395	1.090	.009	.034	-.2114	3.818	.001	.055
	$\hat{\beta}_{cal\text{-}lr}$			.036	.046			.075	.067
	$\hat{\beta}_{naive}$	1.934	3.979	0	.960	2.023	4.3321	0	.977

Table A.6. Simulation Results Logistic-Normal-Gamma Model

			Replication 20%				Replication 5%			
			n = 300							
a,b	σ_w^2	Est.	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)
4,6	36	$\hat{\beta}_{gam}$	-.0010	1.727	.031	.024	-.0003	1.773	.029	.025
		$\hat{\beta}_{norm}$	-.0019	1.863	.035	.023	-.0013	1.983	.035	.024
		$\hat{\beta}_{naive}$	.0084	1.767	.003	.148	.0094	1.899	.003	.160
		$\hat{\beta}_{cal}$	-.0014	1.718	.022	.021	-.0012	1.946	.013	.033
	144	$\hat{\beta}_{gam}$	-.0009	3.289	.027	.020	-.0019	5.318	.036	.018
		$\hat{\beta}_{norm}$	-.0029	4.254	.032	.011	-.0005	6.139	.024	.021
		$\hat{\beta}_{naive}$	.0234	6.103	0	.816	.0243	6.548	0	.860
		$\hat{\beta}_{cal}$	-.0031	4.066	.002	.029	-.0282	996.0	0	.067

a,b	σ_w^2	Est.	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)
1,12	36	$\hat{\beta}_{gam}$	-.0014	1.716	.031	.028	-.0004	1.752	.022	.033
		$\hat{\beta}_{norm}$	-.0037	2.105	.046	.016	-.0028	2.399	.041	.019
		$\hat{\beta}_{naive}$	.0084	1.629	.003	.104	.0083	1.858	.003	.126
		$\hat{\beta}_{cal}$	-.0033	1.955	.025	.019	-.0029	2.360	.018	.026
	144	$\hat{\beta}_{gam}$	-.0035	3.470	.030	.022	-.0048	4.781	.041	.012
		$\hat{\beta}_{norm}$	-.0089	6.941	.070	.011	-.0068	12.00	.046	.014
		$\hat{\beta}_{naive}$	.0211	5.275	0	.656	.0217	5.488	0	.695
		$\hat{\beta}_{cal}$	-.0088	6.096	.007	.022	-.0087	965.0	0	.033

Table A.6. (Continued)

n = 1000

a,b	σ_w^2	Est.	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL<β)
4,6	36	$\hat{\beta}_{gam}$	.0001	.4465	.026	.018	.0002	.5000	.025	.022
		$\hat{\beta}_{norm}$	-.0008	.4811	.034	.018	-.0007	.5552	.035	.013
		$\hat{\beta}_{naive}$	.0089	1.083	0	.384	.0096	1.218	0	.436
		$\hat{\beta}_{cal}$	-.0004	.4465	.024	.020	-.0006	.5372	.020	.024
	144	$\hat{\beta}_{gam}$	.0012	.7804	.020	.021	.0014	.9296	.023	.019
		$\hat{\beta}_{norm}$	-.0001	.9713	.027	.012	.0033	1.314	.025	.012
		$\hat{\beta}_{naive}$	.0236	5.772	0	.999	.0244	6.146	0	.999
		$\hat{\beta}_{cal}$	-.0004	.9077	.013	.040	-.0038	3.209	.001	.055

a,b	σ_w^2	Est.	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)
1,12	36	$\hat{\beta}_{gam}$	-.0003	.4945	.033	.024	.0002	.5028	.029	.027
		$\hat{\beta}_{norm}$	-.0025	.6317	.055	.017	-.0021	.6745	.057	.021
		$\hat{\beta}_{naive}$	.0075	.9115	.003	.254	.0083	1.044	0	.308
		$\hat{\beta}_{cal}$	-.0022	.5850	.045	.018	-.0021	.6554	.032	.027
	144	$\hat{\beta}_{gam}$	-.0014	.8714	.046	.014	-.0008	.9704	.047	.023
		$\hat{\beta}_{norm}$	-.0044	1.412	.083	.004	.0005	1.577	.052	.016
		$\hat{\beta}_{naive}$	.0212	4.742	0	.980	.0244	5.268	0	.994
		$\hat{\beta}_{cal}$	-.0051	1.397	.035	.012	-.0073	4.702	.001	.036

Table A.7. Simulation Results Logistic-t-Normal Model

		Replication 20%				Replication 5%			
				n = 300					
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{nmix}$	-.1943	2.905	.039	.018	-.1294	2.883	.031	.018
	$\hat{\beta}_{norm}$	-.1469	2.724	.032	.023	-.1483	3.044	.029	.017
	$\hat{\beta}_{naive}$	.7180	2.106	.007	.089	.7738	2.074	.003	.088
	$\hat{\beta}_{cal}$	-.1334	2.622	.024	.036	-.1811	3.034	.015	.032
.02	$\hat{\beta}_{nmix}$	-.1639	5.944	.025	.023	.3911	5.778	.020	.038
	$\hat{\beta}_{norm}$	-.2378	7.926	.028	.030	.1385	8.753	.020	.043
	$\hat{\beta}_{naive}$	1.991	4.905	0	.558	2.128	5.445	0	.624
	$\hat{\beta}_{cal}$	-.3027	10.77	.003	.055	-1.289	749.8	0	.085
				n = 1000					
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{nmix}$	.0501	.7603	.033	.040	.0279	.7427	.032	.021
	$\hat{\beta}_{norm}$	.0533	.7482	.028	.040	-.0180	.7702	.033	.025
	$\hat{\beta}_{naive}$	.8228	1.137	.001	.232	.8020	1.076	.001	.236
	$\hat{\beta}_{cal}$	.0565	.7282	.027	.052	-.0386	.7797	.023	.026
.02	$\hat{\beta}_{nmix}$	.1812	1.319	.037	.036	.6128	1.326	.018	.038
	$\hat{\beta}_{norm}$	.1659	1.424	.035	.044	.4614	1.718	.032	.046
	$\hat{\beta}_{naive}$	2.028	3.394	0	.962	2.113	4.710	0	.975
	$\hat{\beta}_{cal}$	.0997	1.393	.015	.060	-.0506	2.980	0	.074

Table A.8. Simulation Results Logistic-t-Normal Model

		Replication 20%				Replication 5%			
				n = 300					
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{nmix}$	-.0808	2.411	.027	.031	-.0439	3.630	.034	.052
	$\hat{\beta}_{norm}$	-.1008	2.380	.027	.029	-.0770	3.383	.034	.049
	$\hat{\beta}_{naive}$	.6631	1.841	.005	.087	.7474	2.010	.007	.104
	$\hat{\beta}_{cal}$	-.0815	2.958	.017	.031	-.1045	6.588	.014	.048
.02	$\hat{\beta}_{nmix}$	.3472	5.625	.020	.131	.7343	8.398	.017	.171
	$\hat{\beta}_{norm}$	-.2973	8.272	.033	.070	.0182	11.97	.034	.070
	$\hat{\beta}_{naive}$	1.794	4.159	0	.460	1.910	4.682	0	.540
	$\hat{\beta}_{cal}$	-.3294	88.81	.003	.079	.9458	830.7	0	.134
				n = 1000					
σ_w^2	Estimate	Bias	MSE	P(UL<β)	P(LL>β)	Bias	MSE	P(UL<β)	P(LL>β)
.005	$\hat{\beta}_{nmix}$	.0095	.9925	.025	.038	.0368	1.260	.031	.040
	$\hat{\beta}_{norm}$	-.0256	1.015	.031	.030	.0017	1.286	.033	.039
	$\hat{\beta}_{naive}$	.7211	.9295	.002	.210	.7883	1.003	.001	.242
	$\hat{\beta}_{cal}$	.0515	14.86	.017	.036	.0723	4.047	.007	.036
.02	$\hat{\beta}_{nmix}$	-.1639	5.944	.025	.023	1.026	2.337	.009	.299
	$\hat{\beta}_{norm}$	-.2378	7.926	.028	.030	.1650	9.375	.061	.109
	$\hat{\beta}_{naive}$	1.991	4.905	0	.558	1.923	4.009	0	.941
	$\hat{\beta}_{cal}$	-.3027	10.77	.003	.055	-3.583	8871	0	.131

Table A.9. Simulation Results Logistic-Lognormal-Lognormal Model

		Replication 20%				Replication 5%			
		n = 300							
σ_w^2	Estimate	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	-.0007	.9120	.033	.022	-.0010	.8913	.035	.017
	$\hat{\beta}_{norm}$	-.0004	.8971	.029	.024	-.0007	.9136	.034	.027
	$\hat{\beta}_{naive}$	.0047	.8978	.011	.102	-.0049	.8320	.005	.092
	$\hat{\beta}_{mcal}$	-.0003	.8426	.025	.034	-.0007	.8812	.027	.043
	$\hat{\beta}_{cal}$	-.0004	.8189	.021	.037	.0001	.8142	.015	.036
.08	$\hat{\beta}_{lognor}$	-.0005	2.440	.025	.035	.0019	4.657	.017	.038
	$\hat{\beta}_{norm}$	.0013	2.933	.023	.057	.0038	6.320	.024	.072
	$\hat{\beta}_{naive}$	.0248	6.479	0	.973	.0258	6.952	0	.990
	$\hat{\beta}_{mcal}$	.0007	2.008	.030	.117	.0006	6.268	.094	.183
	$\hat{\beta}_{cal}$	.0731	4982	.001	.137	.0506	8305	0	.224
		n = 1000							
σ_w^2	Estimate	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)	Bias	MSEx10^{-4}	P(UL<β)	P(LL>β)
.01	$\hat{\beta}_{lognor}$	-.0001	.2332	.024	.020	-.0000	.2391	.021	.020
	$\hat{\beta}_{norm}$	.0002	.2285	.019	.024	.0002	.2491	.021	.025
	$\hat{\beta}_{naive}$	.0051	.4361	0	.230	.0056	.4798	0	.264
	$\hat{\beta}_{mcal}$	.0006	.2201	.016	.036	.0007	.2245	.011	.050
	$\hat{\beta}_{cal}$	.0008	.2173	.012	.033	.0009	.2330	.010	.042
.08	$\hat{\beta}_{lognor}$	.0012	.6498	.030	.036	.0053	1.116	.015	.040
	$\hat{\beta}_{norm}$	.0032	.8272	.023	.079	.0072	1.726	.026	.085
	$\hat{\beta}_{naive}$	.0249	6.308	0	1	.0262	6.939	0	1
	$\hat{\beta}_{mcal}$	.0005	.5352	.024	.124	.0018	1.290	.090	.248
	$\hat{\beta}_{cal}$	.0046	.8327	0	.172	-.0031	285.0	0	.240