ESTIMATION OF QVF MEASUREMENT ERROR MODELS USING EMPIRICAL LIKELIHOOD METHOD

Naima Shifa

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Committee:
Hanfeng Chen, Advisor
Christopher M. Rump
Graduate Faculty Representative
Barbara E. Moses
John T. Chen
Wei Ning
ABSTRACT

Hanfeng Chen, Advisor

Predictor variables are often contaminated with measurement errors in statistical practice. This may be the case due to bad measurement apparatus or just because the true value of the variable cannot be measured precisely. In the framework of general regression models, measurement errors or misclassifications have very serious consequences in many cases as they lead to bias in the estimated parameters that does not disappear as the sample size goes to infinity. In most cases the estimated effect of the contaminated covariate is attenuated. There are some techniques, regression calibration, simulation extrapolation (SIMEX), and the score function method for correcting effect estimates in the presence of measurement error. These widely used approaches have some restricted applications in many situations, for example, SIMEX is a useful tool for correcting effect estimates in the presences of additive measurement error. The method is especially helpful for complex models with a simple measurement error structure. Score function method is employed only for linear measurement error models. In this dissertation, an inference method has been proposed that accounts for the presence of measurement error in the explanatory variables in both linear and nonlinear models. This approach relies on the consideration of the mean and variance function of the observed data and application of the empirical likelihood approach to those functions, which is referred to as quasi likelihood and variance function (QVF). This proposed approach provides the confidence intervals with high inclusion probability of the unknown regression parameters. Moreover, this method is computationally easy to employ to any measurement error model for correcting bias. In addition, general descriptions and comparisons of the existing methods and the suggested
estimation framework with some applications in real life data are discussed. A simulation study is conducted to show the performance of the proposed estimation framework.
ACKNOWLEDGMENTS

I would like to thank my advisor Professor Hanfeng Chen for his continuous support and numerous thoughtful suggestions throughout this research. I also want to extend my gratitude to the other members of my committee, Dr. Barbara Moses, Dr. John Chen, Dr. Christopher Rump, and Dr. Wei Ning for their time and valuable advice. I extend special thanks to all my professors in the Department of Mathematics and Statistics.

I am also grateful to the Department of Mathematics and Statistics at Bowling Green State University for the teaching assistantship and non-service fellowship awarded to me during my study. My very sincere thanks go to Marcia Seubert, Cyndi Patterson, and Mary Busdeker for their assistance.

Finally, my deepest gratitude goes to my parents, and my father and mother in laws. My special thanks to my brother and sister-in-law, Mohona for their sincere support and gratitude to my younger sister for her encouragement, and to my family Mamun, Maieasha, and Naiar for their love, patience, and sacrifice during my study.

Bowling Green, Ohio

Naima Shifa

August, 2009
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Measurement Error ................................................................. 1
   1.1.1 Effects of Measurement Error .............................................. 4

1.2 Linear Errors-in-Variables Models .............................................. 6
   1.2.1 Models for Measurement Error .............................................. 7
   1.2.2 Differential and Nondifferential Error ...................................... 9

1.3 Sources of Data .......................................................................... 10

1.4 Bias Caused by Measurement Error and Correcting for Bias .......... 11
   1.4.1 Method-of-Moments .............................................................. 12
   1.4.2 Orthogonal Regression ......................................................... 13

1.5 A Brief Tour ............................................................................... 13

CHAPTER 2: QUASI-LIKELIHOOD AND VARIANCE FUNCTION (QVF) MODELS 15

2.1 Likelihood Method in Measurement Error Model ............................. 15

2.2 Quasilikelihood and Variance Function (QVF) for Error Model ........ 18
   2.2.1 Description of Methods .......................................................... 19
   2.2.2 Estimation for QVF Models ...................................................... 20

2.3 Idea of Empirical Likelihood ..................................................... 24

2.4 Empirical Likelihood Inference for a Scalar random Variable ........ 31
   2.4.1 Nonparametric Likelihood Ratio ............................................. 33
   2.4.2 Tied Observation ................................................................. 34
   2.4.3 Empirical Likelihood (EL) for a Univariate Mean .................... 34
2.4.4 Coverage Accuracy ................................................................. 35
2.4.5 Power and Efficiency ............................................................... 35
2.4.6 Computational Chores for a Univariate Mean ...................... 36
2.5 Empirical Likelihood for Random Vectors .............................. 37
   2.5.1 Empirical Likelihood for a Multivariate Mean ..................... 38

CHAPTER 3: ESTIMATION OF QVF MEASUREMENT ERROR MODELS USING
   EMPIRICAL LIKELIHOOD METHOD ............................................ 39
   3.1 Overview of regression Calibration ....................................... 40
      3.1.1 The Regression Calibration Algorithm .......................... 43
      3.1.2 The Regression Calibration Estimator for Linear Regression .. 44
      3.1.3 Example of Regression Calibration ................................. 46
   3.2 Overview of Simulation Exploration ..................................... 50
      3.2.1 The Simulation Extrapolation Estimator for Linear Regression ... 52
      3.2.2 The SIMEX Algorithm .................................................. 54
      3.2.3 Example of SIMEX ..................................................... 59
   3.3 Idea of Score Function Method ............................................ 61
      3.3.1 Linear Regression Corrected and Conditional Scores .......... 61
         3.3.1.1 Linear Regression Conditional Scores ..................... 62
         3.3.1.2 Linear Regression Corrected Scores ...................... 63
      3.3.2 Logistic Regression Corrected and Conditional Scores ......... 65
         3.3.2.1 Logistic Regression Conditional Scores .................. 66
         3.3.2.2 Logistic Regression Corrected Scores .................... 66
   3.4 Conditional Score Function: Basic Theory .......................... 67
3.5 Corrected Score Function: Basic Theory ............................................................. 68

3.6 Comparison of Conditional and Corrected Score Functions ......................... 69

3.7 Comparison of Different Approaches Using Real Life Data ............................ 69

   3.7.1 Application of the Method to Simple Linear

       Regression Model .......................................................................................... 70

CHAPTER 4: ESTIMATION OF ERROR-PRONE REGRESSION PARAMETERS USING

       QVF AND EMPIRICAL LIKELIHOOD METHOD ............................................. 75

4.1 Estimation of Regression Parameters .............................................................. 76

   4.1.1 Simple Linear regression Model with Measurement Error ....................... 77

   4.1.2 Estimation of Simple Linear Regression Parameter Using QVF and

       Empirical Likelihood (EL) Method ................................................................. 79

4.2 Multiple Regression: Single Covariate Measured with Error ......................... 82

   4.2.1 Estimation of Multiple Linear Regression Parameter Using QVF and

       Empirical Likelihood (EL) Method ................................................................. 83

   4.2.2 Application of the Method to Simple Linear Regression Model:

       Example ......................................................................................................... 85

4.3 Multiple Regression: Several Covariates Measured with Error ...................... 95

4.4 Estimation of True Values ............................................................................... 98

4.5 Simulation study ............................................................................................. 99

4.6 Estimation of True Values ............................................................................... 102

CHAPTER 5: SUMMARY AND CONCLUSION ......................................................... 105

REFERENCES ..................................................................................................... 107

APPENDIX ......................................................................................................... 117
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Total Somites</td>
</tr>
<tr>
<td>3.1</td>
<td>Index of the Number of Hen Pheasants in Iowa</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparison between the Naïve and Regression Calibration Estimators</td>
</tr>
<tr>
<td>3.3</td>
<td>Comparison between the Naïve and SIMEX Estimators</td>
</tr>
<tr>
<td>3.4</td>
<td>Yield of Corn on Marshall Soil in Iowa</td>
</tr>
<tr>
<td>3.5</td>
<td>Comparison of estimators: Naïve, Regression Calibration, SIMEX for the Simple Linear Regression Model with Single Variable Measured with Error</td>
</tr>
<tr>
<td>4.1</td>
<td>Data from role performance study</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of Estimators: Naïve, Regression Calibration, SIMEX and QVF and EL Method for the Linear Regression Models with Single Variable Measured with Error</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison of Estimators: Naïve, Regression Calibration, SIMEX and QVF and EL Method for the Multiple Linear Regression Models with Several Variables Measured with Error</td>
</tr>
<tr>
<td>4.4</td>
<td>Simulation Output</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Histogram of the Number of Somites of 487 Earthworms Reported in (1905)</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>Scatter plot of Spring hens versus August hens</td>
<td>49</td>
</tr>
<tr>
<td>3.2</td>
<td>Scatter plot of soil nitrogen and yield of corn</td>
<td>72</td>
</tr>
<tr>
<td>4.1</td>
<td>Scatter plot of Knowledge and Role Performance</td>
<td>90</td>
</tr>
<tr>
<td>4.2</td>
<td>Scatter plot of Value Orientation and Role Performance</td>
<td>91</td>
</tr>
<tr>
<td>4.3</td>
<td>Scatter plot of Role Satisfaction and Role Performance</td>
<td>92</td>
</tr>
<tr>
<td>4.4</td>
<td>Scatter plot of Past Training and Role Performance</td>
<td>93</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Problems of measurement error spread through all research areas; from agriculture to astronomy. A common source of measurement error problem comes from incorrect response to a survey question, incorrect coding of a correct response and the use of a correctly measured variable as a proxy for another theoretically valid but unobserved variable. Questions that seek sensitive information may elicit partial or incorrect responses. That is measurement error is triggered by unobservable when such variables are replaced by proxy variables. In the first section of this chapter we introduce the idea of measurement error problems and investigate the consequences of measurement errors on the least square estimators. Furthermore, simple linear regression model is considered for this purpose. In the next two sections we outline the brief discussions of different types of measurement error models with examples. We also describe the sources of the data and the bias caused by measurement error and correcting for bias. In the last section we bring in the brief plan of the later chapters.

1.1 Measurement Error

Broadly speaking the consequence of errors of measurement is a failure to identify the parameters of interest. The issue of fixing the problem is complex. One may consider simply omitting the relevant variable in the model or substituting a proxy for the true measure. There are at least two important reasons for not doing so except in extreme cases. First if the variable is of central interest, then omission leads to serious omitted variable bias, so one is replacing one type of problem for another, and identification is still not possible. Second, in a linear regression, using a proxy for a latent variable will have smaller asymptotic bias than simply omitting the variable from the model. Ignoring the variable provides inferior estimates and using the proxy still gives inconsistent estimates even though the biases are smaller. The essential insight
underlying the solution of the measurement error problem is that to recover the parameter of the unobserved variable and to identify the model. It is necessary to have extraneous information in the form of additional assumptions about the measurement error or obtain additional data and use this information after invoking plausible assumptions.

Measurement error have potentially very serious consequences since in many cases they lead to regression parameters becoming unidentified. The precise consequences of measurement errors may depend on the functional form of the model, how the errors enter the model, and the data structure under consideration. The solution of the problem resulting from measurement errors typically requires introduction of additional information into the model, either in the form of additional data or additional assumptions. Many scientific studies involve fitting a relationship of the form

\[ Y = g(X; \theta) + \varepsilon \]  

Here \( Y \in IR \) is the response variable, \( X \in IR \) is the predictor variables, \( \theta \in IR^p \) is the unknown regression parameter and \( \varepsilon \) is the random error. In many applications, statistically meaningful models are defined in terms of predictor variables \( X \) that for some reason cannot be measured directly, or it is measured with substantial random error. In such situation it is common for substitute variables \( W \) to be observed instead. The substitution of \( W \) complicates the statistical analysis of the observed data when the purpose of the analysis is inference about a model defined in terms of \( X \). The problems of this nature are commonly called measurement error problems and the statistical models and methods for analyzing such data are called measurement error models. For example, we can consider the relationship between the yield of corn and available nitrogen in the soil. Assume that (1.1.2) is an approximate relationship between the yield and nitrogen.
\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n, \]

(1.1.2)

The coefficient \( \beta_1 \) is the amount that the yield is increased or decreased when soil nitrogen increases or decreases one unit. But we cannot observe the exact amount of soil nitrogen \( X \). To estimate the available soil nitrogen we need to sample the soil of the experimental plot and perform a laboratory experiment. Because of sampling and laboratory experiment measurement error is introduced. Similarly, a person’s measured systolic blood pressure \( W \) differs from his or her long-term average systolic blood pressure \( X \) because of significant temporal variation as well as instrument and reader error. Again, one can think about the study of the relationship between the temperature used to dry a sample for chemical analysis and the resulting concentration of a volatile constituent, an oven is used to prepare samples. The temperature \( X \) is set at 250, 350, 400, 450 and 550 degrees Fahrenheit, respectively. The true temperature inside the oven, however, may vary randomly around the pre-set values. In such a situation (1.1.1) is a reasonable model for the measurement error. The measurement error models are extensively used in nutrition and epidemiology. Consider the study relating radiation exposure to disease, like the Nevada Test Side (NTS), described by Stevens, Till, Thomas, et al. (1992); Kerber, Till, Simon, et al. (1993); and Simon, Till, Lloyd, et al. (1995) where the radiation exposure largely came as the result of above-ground nuclear testing in 1950s. In this study more than 2,000 individuals were exposed to radiation and children were examined for the thyroid disease. The primary exposure came from milk and vegetables. The plan of the study was to relate various thyroid disease outcomes to radiation exposure to the thyroid. The main exposure of interest, radiation to the thyroid, cannot be observed exactly. A model is considered to convert the known data about the above-ground nuclear tests to radiation actually absorbed into the thyroid. Calculations were based on age at exposure, gender, residence, history, x-ray history, whether...
the individual was a breast-fed child, and a diet questionnaire filled out by the parent, focusing on milk consumption and vegetables. The researchers put these data into a model and find the point estimate of thyroid dose and associated standard error. The measurement error or uncertainty was taken into account in dose estimate.

1.1.1 Effects of Measurement Error

The effects of measurement error have long been recognized. Measurement error in the regressors is an important topic as it leads to inconsistency of the ordinary least square estimator even if the measurement error has zero mean. Measurement error in the regressors is often said to lead to bias, but actually it leads inconsistency, as the bias does not disappear as the sample size goes to infinity. For the standard regression model (1.1.2) with single explanatory variable, $X_i$ measured with error or not observed directly, we assume that the $X_i$s are independently drawing from a $N(\mu_X, \sigma_{XX})$ distribution. It is also assumed that the vector $(\varepsilon_1, \ldots, \varepsilon_n)$ is independent of the vector $(X_1, \ldots, X_n)$. Instead of observing $X_i$, we observe the sum

$$W_i = X_i + U_i$$

Here, $U_i$ represents measurement error. To investigate the effect of measurement error on the least square coefficient consider the simple models (1.1.2) and (1.1.3), under the assumption

$$\begin{bmatrix} 1 & X_i & U_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \mu_X \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{XX} & \sigma_{XU} & \sigma_{U} \\ \sigma_{XU} & \sigma_{UU} & \sigma_{U} \\ \sigma_{U} & \sigma_{U} & \sigma_{U}^2 \end{bmatrix}\right)$$

The vector $(Y_i, X_i)$ is distributed as bivariate normal with mean (1.1.5a) and variance (1.1.5b)

$$E\{(Y, W)\} = (\mu_Y, \mu_W) = (\beta_0 + \beta_1 \mu_X, \mu_X)$$

$$\begin{bmatrix} \sigma_{YY} & \sigma_{YW} \\ \sigma_{Wy} & \sigma_{WW} \end{bmatrix} = \begin{bmatrix} \beta_1^2 \sigma_{XX} + \sigma_{xx} & \beta_1 \sigma_{XX} \\ \beta_1 \sigma_{XX} & \sigma_{XX} + \sigma_U \end{bmatrix}$$

Let $\hat{\beta}_{Y|W}$ denote the slope estimator from the least squares regression of $Y$ on $W$ defined as
\[ \hat{\beta}_{Y|W} = \left( \sum_{i=1}^{n} (W_i - \bar{W})^2 \right)^{-1} \sum_{i=1}^{n} (W_i - \bar{W})(Y_i - \bar{Y}) \]  

(1.1.6)

By the properties of the bivariate normal distribution,

\[ E \left\{ \hat{\beta}_{Y|W} \right\} = \sigma_{ww}^{-1} \sigma_{wy} = \beta_1(\sigma_{xx} + \sigma_U^{-1}) \sigma_{xx} = \beta_1 \]  

(1.1.7)

Similarly let \( \hat{\beta}_{Y|X} \) denote the slope estimator obtained by least squares regression of \( Y \) on \( X \).

Then for the classical model with fixed \( X \), the estimator,

\[ \hat{\beta}_{Y|X} = \left( \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \]  

(1.1.8)

is called a linear estimator. Since the measurement errors \( U_i \) have mean \( \mu_U = 0 \) and constant variance \( \sigma_U \) and are independent of the true predictors \( X_i \) and the equation error \( \epsilon_i \), it transpires that

\[ \hat{\beta}_{Y|W} = \hat{\rho} \hat{\beta}_{Y|X} + o_p(1) \]  

(1.1.9)

Here

\[ \hat{\rho} = \frac{s_x^2}{s_x^2 + s_U^2} \]  

(1.1.10)

The notation \( s^2 \) denotes the sample variance of the subscript variable. Because \( 0 < \hat{\rho} < 1 \), a consequence of ignoring measurement error is attenuation in the slope estimator, that is, bias towards zero. If there are errors of measurement in the explanatory variable, the error in the least square estimator is not linear in the full set of random errors. Considering (1.1.3), the error in the least square estimator is
\[
\hat{\beta}_{i, W} - \beta_i = \left[ \sum_{i=1}^{n} \left( (X_i - \bar{X})^2 + 2(\bar{X} - \bar{U})(\bar{U} - \bar{U}) + (\bar{U} - \bar{U})^2 \right) \right]^{-1} \\
\times \left[ \sum_{i=1}^{n} \left( (X_i - \bar{X})(v_i - \bar{v}) + (\bar{U} - \bar{U})(v_i - \bar{v}) \right) \right],
\]

here \( v_i = \varepsilon_i - U_i \beta_i \)

In short measurement error in explanatory variables has some adverse affects, like:

- It brings in bias in parameter estimation for statistical models.
- It causes a loss of power in detecting the relationship among the variables.
- It masks the features of the data and makes it difficult to analysis the graphical models.

1.2 Linear Errors-in-Variables Models

Generalizations and variations of linear error-in-variables are extensively studied; generalizations include multiple and multivariate multiple linear regression, in which one or more predictor variables are measured with error and some predictor variables are error free. The variations depend on assumptions about the unobserved predictors and the type of data or distributional information available, ensuring identifiably of the regression parameters.

In statistical literature, it is conventional to distinguish between functional and structural approaches to measurement error models. In *Functional errors-in-variable models*, the unobserved \( X_i \) are modeled as unknowns, nonrandom constants, but in *Structural errors-in-variable models*, the observables and unobservables jointly vary in repeated sampling. The observed variable \( W_i \) is called the *manifest variable* or the *indicator variable*. The unobserved variable \( X_i \) is called a latent variable in certain areas of application. That is, models with fixed \( X_i \) are called *functional models*, while models with random \( X_i \) are called *structural models*. For example, consider the relationship between aquatic species diversity \( Y \) and acid neutralizing
capacity \( X \), given measurements of \( (Y, X) \) from each of \( n \) lakes. If the only lakes of interest are those represented in the sample, then it is appropriate to model \( X_i, i = 1, \ldots, n \), as unknown constants. On the other hand, if the lakes represented in the data are a random sample from a large population of lakes, then it is appropriate to model \( X_i, i = 1, \ldots, n \), as independently, identically distributed random variables. Again, if we consider soil nitrogen data and think about the fields are a set of experimental fields managed by the experiment station in ways that produce different levels of soil nitrogen in the different fields. This can be done by the varying rates of fertilizer and growing of different fractions of legumes in the rotations would produce different levels of soil nitrogen. In such a situation, we could treat the true, but unknown, nitrogen levels in the different fields as fixed. Again, if the fields are a random sample of farmers’ fields, the true values of soil nitrogen could be treated as random variables.

1.2.1 Models for Measurement Error

Classical measurement errors arise when a quantity is measured by some device and repeated measurements vary around the true value. In classical measurement error model the truth is measured with additive error, usually with constant variance. For example, in cancer study if someone is interested in measuring the logarithm of dietary protein intake, the true long-term log-intake cannot be observed in practice. Instead, the investigators measure a biomarker of log-protein intake, namely urinary nitrogen measurements. In this study there is evidence that the protein biomarker captures true protein intake with added variability. In symbols, let \( X_i \) be the true log-protein intake for individual \( i \), and let \( W_{ij} \) be the \( j^{th} \) biomarker log-protein measurement. Then the classical measurement error model states that

\[
W_{ij} = X_i + U_{ij}
\]  

(1.2.1)
In this model, \( W_i \) is an unbiased measure of \( X_i \), so that \( E(U_i \mid X_i) = 0 \) and the error structure of \( U_i \) could be homoscedastic or heteroscedastic. In this model, the variability of the observed log protein-intake is greater than the variability of the true protein intake.

Unlike classical error model, Berkson measurement error model causes little bias in the measurement. In the Berkson error model, the observed values \( W \)'s are fixed in repeated sampling and the \( X \) values vary; it is extensively studied in nonlinear measurement error models. Consider an experiment in which the quality of cement, \( Y \), is to be studied as a function of the amount of water in the mixture. The amount of water is controlled by setting a metered valve to specified values. Because of fluctuation in water pressure and inaccuracies in the metered valve, the true amount of water in a mixture \( X \) differs from the prescribed amount \( W \). Here \( W \) is controlled as part of the experimental design and is not random. In replication of the experiment at the same value of \( W \), the true amount \( X \) will vary. If the value is correctly calibrated, then a reasonable model is,

\[
X_i = W_i + U_i
\]

Moreover, \( E(U_i \mid W_i) = 0 \) regardless of the value of \( W \). Accordingly, for unbiased Berkson error, \( E(X \mid W) = W \); and for unbiased classical error, \( E(W \mid X) = X \). In short if the true dose has more variability than the estimated dose, Berkson measurement error model is proposed for it. It is assumed that the true dose is equal to the estimated dose plus measurement error.

Application of Berkson error models is common in epidemiological studies. For example, a ubiquitous radioactive gas, radon is the second leading cause of lung cancer after smoking in general population (NAS, 1994). Epidemiological studies on lung cancer and residential radon exposure have been studied in many countries to obtain relative risk (RR) estimates, and to describe the exposure-disease relationship. In Western and Eastern parts of Germany two case-
control studies were conducted in the 1990s consisted of 2500 patients diagnosed with lung cancer from hospitals and a group of 4000 disease-free participants recruited by population registry. Data were collected on their long-term residence history, smoking status, and occupational history. Alpha track detectors measured the radon concentration in the bedroom and the living room of their homes over one year. Based on these measurements and information on the time of the rooms’ occupancy and on each home residency, the radon exposure was assessed retrospectively.

Measurement errors in exposure assessment are unavoidable, with residential radon exposure being no exception, induce bias on RR estimates. The true radon gas concentration in an environment is the concentration of radon in a certain time point; but what is actually measured is the average concentration during exposure of the detector. Berkson type error model is appropriate in this situation since a group’s average is the “measured value”, that is the value that enters the analysis, and individual latent value is the “true value”. Actually, Berkson errors include the use of job-exposure-matrix (JEM), which is a cross-classification of occupations and workplace exposures over time, instead of individual exposure measurements or the use of environmental exposure measurements via fixed monitors instead of individual dose measured via personal dosimeters.

There is a major difference between Berkson and classical measurement error when one is attempting power calculations.

1.2.2 Differential and Nondifferential Error

Nondifferential measurement error occurs when the observed variable or surrogate, \( W \) contains no information about the response, \( Y \) other than what is available in \( X \) and \( Z \). That is the measurement error is nondifferential if the distribution of \( Y \) given \( (X, Z, W) \) depends only
on \( (X, Z) \). In other words, \( W \) is conditionally independent of the response given the true covariates; otherwise the measurement error is \textit{differential}. For example, someone is interested in long-term systolic blood pressure, \( X \) but we can only observe blood pressure on a single day, \( W \). A single day’s blood pressure contributes essentially no information about long-term blood pressure, and hence the measurement error is nondifferential. Many problems have been classified as having nondifferential measurement error, when the true and observed covariates occur at a fixed point in time and the response is measured at a later time.

Differential measurement error occurs when the response is obtained first and then subsequent follow-up ascertains the covariates or when the surrogate, \( W \) is not merely a mismatched version of \( X \), but is a separate variable acting as a type of proxy for \( X \). For example, a woman who develops breast cancer may well change her diet, so the reported diet measured after the diagnosis is correlated with cancer outcomes, even after taking into account long-term diet before diagnosis.

1.3 Sources of Data

Measurement error is a serious problem nearly all fields of science. There are many possible sources of error, such as use of inexact measure because of high costs of exact evaluation, use of substitutes for latent variables that cannot be measured at all, or imperfect recall by study subjects. Measurement errors can have mean zero or systematically enlarge or decrease observed values of a variable in relation to their true values. Examples include technical mistakes, like, systematic rounding errors or systematic mistakes in coding of responses, or answers to ‘sensitive’ questions, in which respondents tend to palliate their answers. Some error sources can be reduced in advance by means such as careful planning, rigorous training of interviewers, or validation of questionnaires. But some error will often remain. For measurement error data
analysis, one must need information about either $W$ given $(X, Z)$ or about $X$ given $(Z, W)$.

There are two main sources of measurement error data, internal subsets of the primary data and external or independent studies (Carroll, et al, 2006); there are also three types of data within these broad categories, validation data, replication data, and instrumental data. In validation data $X$ is obtained directly but subject to measurement error. In this case measurement error is an imprecise measurement of exact figure, giving some information on the true value. For example, rounding of income figures may be regarded as a measurement error (Hanisch 2005, 2006). In measurement error analysis, one tries to establish the knowledge of what has not been precisely observed. In replication data, the replicates of $W$ are available. Another variable $T$ is observable with $W$ in instrumental data. An instrumental data set is ideal for all known statistical techniques and permits direct examination of the error structure and leads to greater precision of estimation and inference. External validation data can also be used in analyzing the error model with an assumption when transporting such models to primary data. If replicated mean is a better estimate of $X$ than a single observation, then replicated measurements can be made. Replication data can be used to estimate the variance of the measurement error $U$. In the data set sometimes there is a second measurement $T$ with the primary measurement $W$, which may or may not be unbiased for $X$. If $T$ is internal, it is called instrumental variable and it can be used in instrumental variable analysis provided that $T$ possesses certain other statistical properties. If $T$ is external, it can be used in regression calibration analysis.

1.4 Bias Caused by Measurement Error and Correcting for Bias

The effect of measurement error depends on the model under consideration and on the joint distribution of the measurement error and other variables; it also depends whether or not the error prone variable is univariate or multivariate and the presence of bias in the measurement.
Measurement error causes attenuation and hides the real effects and exhibits the relationships that are not present in the error-free data. Sometimes it even reverses the signs of the estimated coefficients. Since the measurement error distribution determines the effect of the measurement error, appropriate methods for correcting the effects of measurement error depend on the measurement error distribution. Methods-of-moments and orthogonal regressions are two common methods for eliminating bias.

1.4.1 Method-of-Moments

The method-of-moments estimators can be constructed for general linear regression model,

\[ Y = \beta_0 + \beta_x^t X + \beta_z^t Z + \varepsilon \]  

(1.4.1)

There are covariates \( Z \) measured without error and \( W \) is unbiased for \( X \), the multiple predictors.

Then the na"ive ordinary linear regression consistently estimates not \( (\beta_x \beta_z) \) but \( (\beta_{x*} \beta_{z*}) \); see Carroll, Ruppert, et al. (2006)

\[
(\beta_{x*} \beta_{z*}) = \left( \sum_{xx} + \sum_{UU} \quad \sum_{xz} \right)^{-1} \left( \begin{pmatrix} \sum_{xy} \\ \sum_{zy} \end{pmatrix} + \begin{pmatrix} \sum_{Ue} \\ 0 \end{pmatrix} \right)
\]

(1.4.2)

If \( \sum_{UU} \) and \( \sum_{Ue} \) are known or can be estimated, (1.4.2) can be used to construct a simple method-of-moments estimator to correct bias. The method-of-moments estimator that corrects for the bias in the case when \( \sum_{UU} \) and \( \sum_{Ue} \) are known is,

\[
\begin{pmatrix}
S_{ww} + \sum_{UU} & S_{wz} \\
S_{zw} & S_{zz}
\end{pmatrix}^{-1} \begin{pmatrix}
S_{wy} - \sum_{Ue} \\
S_{zy}
\end{pmatrix}
\]

(1.4.3)
If $\Sigma_{UU}$ and $\Sigma_{U_e}$ are unknown, then the estimated values are used in (1.4.3) to correct bias. It is often reasonable to assume that $\Sigma_{U_e} = 0$ which simplifies (1.4.3).

### 1.4.2 Orthogonal Regression

Like the method-of-moments estimator, orthogonal regression requires the knowledge of measurement error variance, $\sigma_U^2$ and $\eta = \frac{\sigma_{ee}}{\sigma_U^2}$, where the linear regression model is,

$$Y = \beta_0 + \beta_x X + \varepsilon$$

(1.4.4)

$$W = X + U$$

(1.4.5)

Both $\varepsilon$ and $U$ are uncorrelated. The orthogonal regression estimator minimizes the orthogonal distance of $(Y, W)$ to the line $\beta_0 + \beta_x X$ weighted by $\eta$, that is it minimizes,

$$\sum_{i=1}^{n} \left( (Y_i - \beta_0 - \beta_x X_i)^2 + \eta (W_i - X_i)^2 \right)$$

(1.4.6)

in unknown parameters, $(\beta_0, \beta_x, X_1, \ldots, X_n)$.

The orthogonal regression estimator is the functional maximum likelihood estimator assuming $(X_1, \ldots, X_n)$ are unknown fixed constants and the errors $(\varepsilon, U)$ are independent and normally distributed. Orthogonal regression has greater applicability than method-of-moments estimation in that only the ratio $\eta$, of the error variances need to be known or estimated. The value of $\eta$ should be properly specified, otherwise it results an unacceptable large value over correction for attenuation.

### 1.5 A Brief Tour

This work is structured into three parts; the first part consists of first two chapters. In here the readers may find an idea about the measurement error problem and the reasons for fixing it. Since the empirical likelihood method is extremely useful for constructing confidence intervals
or regions of the parameter of interest and has been extensively applied to linear regression and generalized linear regression models; the purpose of this is to show this necessity with a real life example. Moreover, this part includes the empirical likelihood inference for univariate and multivariate means. The second part of this work presents the existing methods that extensively used for the measurement error problems. The last part describes a new approach, which is simply based on the first two moments (QVF) and the empirical likelihood approach to make inference of regression parameters of the error prone predictors. In addition, this part shows the comparisons of the existing methods and new approach. For this comparison data sets from real life are considered. Furthermore, a simulation study was conducted to show the performance of the proposed method.
CHAPTER 2
QUASI-LIKELIHOOD AND VARIANCE FUNCTION (QVF) MODELS

Likelihood approach is perhaps the most important concept for inference in parametric models. Recently it has also been shown to be helpful in nonparametric contexts. The first section of this chapter demonstrates the idea of the likelihood method in measurement error perspectives with the description of quasi-likelihood function which is a function with similar properties to the log-likelihood function, except that function is not the log-likelihood corresponding to any actual probability distribution. Quasi-likelihood models can be fitted using a straightforward extension of the algorithms used to fit generalized linear models. The second section describes quasi-likelihood and variance function models for measurement error with the derivations. In this model instead of specifying a probability distribution for the data, only a relationship between the mean and the variance is specified in the form of a variance function giving the variance as a function of the mean. Importance of empirical likelihood method is postulated in the third section with example and some applications.

2.1 Likelihood Method in Measurement Error Model

The purpose of the likelihood function is to convey information about unknown quantities. The unknown quantities may be fixed parameter with associated estimation problem, or unobserved random values; in a prediction problem the two unknowns can easily be mixed. The likelihood provides a measure of relative preference of various parameter values. If it is assumed that a statistical model is parameterized by a fixed unknown \( \theta \), the likelihood \( L(\theta) \) is the probability of the observed data \( X \) considered as a function of \( \theta \). The data \( X \) including any set of observations from an experiment and the parameter \( \theta \) can be a vector of values. The likelihood method requires stronger distributional assumptions, but they can be applied to more
general problems. The likelihood for a fully specified parametric distribution can be applied to find likelihood ratio confidence intervals. Likelihood based confidence intervals are more reliable in nonlinear problems than for the normal approximation or the bootstrapping. In the context of the measurement error problem, advantages of likelihood method relative to functional method have been studied by Schafer and Purdy (1996) and Küchenhoff and Carroll (1997). The advantageous properties are contingent on the correct specification of the likelihood. Sometimes likelihood method is computationally more demanding and often difficult to understand modeling assumptions due to robustness. To perform likelihood analysis in measurement error one must define a parametric model for every component of the data. This analysis begins with an outcome model for the distribution of the response given the true predictor. The likelihood of an observed data point \((Y, W)\) conditional on \(Z\) is, 

\[
 f_{Y|X,Z}(y | x, z, \beta)
\]

and we are interested in estimating \(\beta\). The likelihood for the observed data is,

\[
 \prod_{i=1}^{n} f_{Y|X,Z}(y_i | x_i, z_i) \quad \text{and the maximum likelihood estimates are obtained by maximizing the likelihood over all the unknown parameters.}
\]

If the response \(Y\) follows a normal distribution with mean \(\beta_0 + \beta_1' X + \beta_2' Z\) and variance \(\sigma^2\), then 

\[
 f_{Y|X,Z}(y | x, z, \beta) = \sigma^{-1} \phi \left( \frac{y - \left(\beta_0 + \beta_1' x + \beta_2' z\right)}{\sigma} \right),
\]

here \(\phi(\cdot)\) is the density of the standard normal distribution. If the response \(Y\) follows a logistic regression model with mean 

\[
 \mu_{Y|X,Z}(\beta_0 + \beta_1' x + \beta_2' z),
\]

then the likelihood function becomes,

\[
 f_{Y|X,Z}(y | x, z, \beta) = \left[ \mu_{Y|X,Z}(\beta_0 + \beta_1' x + \beta_2' z) \right]^{y} \times \left[ 1 - \mu_{Y|X,Z}(\beta_0 + \beta_1' x + \beta_2' z) \right]^{1-y}
\]
After specifying a parametric model, one must choose an appropriate measurement error model, classical measurement error model, or a Berkson error model, or a combination of any two. If all variables, response $Y$, surrogate $W$, and error-prone predictor $X$ are discrete, then

$$\Pr \left[ Y = y, W = w \right] = \sum_x \Pr \left[ Y = y, W = w, X = x \right]$$

$$= \sum_x \Pr \left[ Y = y \mid W = w, X = x \right] \times \Pr \left[ W = w, X = x \right] \quad (2.1.1a)$$

Since $W$ is surrogate and gives no information about the response $Y$ when $X$ is known, we have,

$$\Pr \left[ Y = y, W = w \right] = \sum_x \Pr \left[ Y = y \mid X = x, \beta \right] \times \Pr \left[ W = w \mid X = x \right] \times \Pr \left[ X = x \right] \quad (2.1.1b)$$

In the case of the classical measurement error model, with no $Z$, one should specify a model for $W$ given $X$ and then a model for $X$, that is,

$$\Pr \left[ Y = y, W = w \right] = \sum_x \Pr \left[ Y = y \mid X = x, \beta \right] \times \Pr \left[ W = w \mid X = x \right] \times \Pr \left[ X = x \right] \quad (2.1.1c)$$

For Berkson type error model, the form of the probability becomes,

$$\Pr \left[ Y = y, W = w \right] = \sum_x \Pr \left[ Y = y \mid X = x, \beta \right] \times \Pr \left[ X = x \mid W = w \right] \times \Pr \left[ W = w \right] \quad (2.1.1d)$$

Since $W$ is surrogate and contains no additional information about the parameter $\beta$, then we have,

$$\Pr \left[ Y = y \mid W = w \right] = \sum_x \Pr \left[ Y = y \mid X = x, \beta \right] \times \Pr \left[ X = x \mid W = w \right] \quad (2.1.1e)$$

To obtain the estimates of the unknown parameters and make the inference one must compute the likelihood function. Since $X$ is unobservable, sometimes it is difficult and time consuming to obtain such a function. To find the estimate of the parameters one takes the product over the
sample of size, $n$ and maximizes the logarithm of the overall likelihood in the unknown parameters. The likelihood function can be maximized by numerical optimization techniques, or by missing data techniques; see Little and Rubin (2002), Tanner (1993) and Geyer and Thompson (1992). Algorithms for computation and maximization of the likelihood in general regression models with measurement error are studied in Higdon and Schafer (2001) and Schafer (2002).

2.2 Quasi-likelihood and Variance Function (QVF) for Error Model

General linear regression models are widely used in the modern regression analysis. Successful modeling based on generalized linear models relies on correctly specified model components including the random part and the systematic part. Random parts require specification of a distribution from the exponential family. The quasi-likelihood function proposed by Wedderburn’s (1974) broadened the scope of generalized linear models by specifying the variance function instead of the entire distribution. To study the quasi-likelihood function one requires assumption on the first two moments only. The advantage of this approach is that no probability structure has to be specified as the estimating function is constructed from the first two moments only.

It is a very useful approach because in many situations the exact distribution of the observations is unknown and this function has statistical properties similar to those of a log-likelihood function. The original quasi-likelihood function cannot be used to make inference about free parameters contained in a variance function. To fix this quasi-likelihood and variance function (QVF) model has been proposed (Mc Cullagh and Nelder, 1989; Carroll, Ruppert and Stefanski, 1995). A model in the quasi-likelihood and variance function framework requires only the knowledge of the relationship between the mean and variance. The QVF model is considered
when the predictor X is measured with error and the surrogate W is observed in QVF model analysis; for a primary data set (Y, W), a validation data set exists for which (X, W) is observed. The QVF model for the response over the predictors X and Z has the following functional structure

\[ E(Y \mid X, Z) = \mu(X, Z, \beta) \]  \hspace{1cm} (2.2.1a)

\[ \text{Var}(Y \mid X, Z) = \sigma^2 v^2(X, Z, \beta, \theta) \]  \hspace{1cm} (2.2.1b)

Here, \( \mu(\cdot) \) is the mean function, \( \beta \) is the vector of the regression parameters, and \( v(\cdot) \) is the variance function and \( \theta \) and \( \sigma^2 \) are called the variance function parameters. It is convenient to separate the variance function parameters into the scale factor \( \sigma^2 \) and \( \theta \) which determine the possible heteroscedasticity. Let \( k \) be the number of dimensions of \( \beta \) and suppose a subset, \( \beta^* \) of \( r \) components of \( \beta \) is of interest in study. Based on the observed covariates and surrogates, an estimate and confidence region for \( \beta^* \) is obtained. This model is widely applicable; including generalized linear models and many important nonlinear regression models. The key feature of (2.2.1a) and (2.2.1b) is that specification of the mean and variance function is sufficient to obtain estimates of the parameters \( \beta \), \( \theta \) and \( \sigma^2 \); it is not needed to specify a likelihood function. Models (2.2.1a) and (2.2.1b) include examples from epidemiology, econometrics, fisheries research, quality control, pharmacokinetic, assay development, etc.

2.2.1 Description of Methods

The QVF model analysis follows the same rules as likelihood analysis, and the necessary steps are mentioned as follows:

Step 1: Identify the mean and variance of Y if X were observed.
Step 2: State a model relating \( W \) to \( X \); that is a model for \( W \) given \( (X, Z) \) and a model for \( X \) given \( Z \).

Step 3: Obtain the estimates of the nuisance parameters.

Step 4: Find the density or mass function of \( X \) given \( (Z, W) \) and use it to obtain (2.2.1a) and (2.2.1b) by integration. The sandwich method or bootstrap can be used for inference, and we apply the empirical likelihood method for this purpose.

### 2.2.2 Estimations for QVF Models

In many circumstances, instead of observing \( X \), a surrogate \( W \) is observed. Ignoring the measurement error and simply substituting \( W \) for \( X \) is inappropriate because it will lead inconsistent parameter estimates. Estimations of unknown parameters based on observed \( (Y, W) \) are desirable. Under the assumption of \( X \) being given, \( W \) is independent of \( Y \), and the mean and the variance function of the observed data are defined as,

\[
E \left( y_i | w_i, z_i \right) = \mu(\hat{x}_i, z_i, \beta) \tag{2.2.2a}
\]

\[
Var \left( y_i | w_i, z_i \right) = \sigma^2 \nu^2(\hat{x}_i, z_i, \beta, \theta) \tag{2.2.2b}
\]

The above equations show that the distribution of \( X \) given \( W \) plays an important role in the estimation of unknown parameters. If the distribution of \( X \) given \( W \) is known up to a finite set of parameters, (2.2.2a) and (2.2.2b) become as parametric formula for the mean and variance functions of the observed data \( (Y, W) \).

The measurement error estimation problem has been discussed by Carroll et. al (1984), Fuller (1987) and Schafer (1987) when a model of \( X \) given \( W \) is postulated. Stefanski (1985), Stefanski and Carroll (1985), Whittemore and Keller (1988) and Carroll and Stefanski (1990)
considered small error approximation to construct estimates adjusted for measurement error. Rosner et al (1989, 1990), Whittemore (1989), Gleser (1990), Pierce et al (1990) and Carroll and Stefanski (1990) discussed the possibilities of directly replacing $X$ in (2.2.2a) and (2.2.2b) by a linear regression for $(X \mid W)$.

In QVF model no assumption is needed about the distribution of $X$ given $W$. In some situations, an independent validation data set, like $(Y,X,W)$ or $(X,W)$ is available; (see Roner et al (1989) and Pepe and Flemming (1991)) for this case Carroll and Wand (1991) and Pepe and Flemming (1991) developed the estimated density of $(X,W)$ and therefore the density of $Y$ given $W$. Pepe and Flemming (1991) proposed to estimate the probability density function of $Y$ given $W$ by its empirical distribution for discrete $W$. Carroll and Wand (1991) discussed the logistic regression and suggested the kernel regression to estimate $\Pr[Y = 1 \mid W = w]$ from validation data. The use of kernel regression for the estimation of the density of $X$ given $W$ simplifies the calculations. Stefanski and Carroll (1993) discussed a new approach to estimate $E(Y \mid W = w)$ and variance, $Var(Y \mid W = w)$ as functions of $(\beta, \theta, \sigma)$ and $w$ by means of kernel regression applied to the validation data. They applied the quasi-likelihood variance function estimation techniques to the primary data using the estimated mean and function, which requires kernel regression estimation of functions of $X$ regressed on $W$ in the validation data. Sometimes the range of $W$ in the observed validation data is smaller than that in the observed primary data. If the kernel technique is used blindly, it would lead extrapolation and would not provide good estimates.
Specification of the mean and variance models (2.2.1a) and (2.2.1b) allow constructing estimates of the unknown parameters in the model. Given $\theta$, $\beta$ can be estimated by generalized least square (GLS) estimator which minimizes $\sum_{i=1}^{n} \{Y - \mu(X, Z, \beta)\}^2$ by solving

$$
\sum_{i=1}^{n} \{Y - \mu(X, Z, \beta)\} \mu'(X, Z, \beta) = 0 \quad (2.2.3)
$$

where $\mu'(X, Z, \beta)$ is the vector of the partial derivatives of the mean function $\mu(X, Z, \beta)$. The least square estimator is inefficient and can be improved by weighting the sum (2.2.3) by reciprocal variance. It is a conditionally unbiased estimating function for estimating $\beta$ which is defined as,

$$
\frac{Y - \mu(X, Z, \beta)}{\sigma^2 v^2(X, Z, \beta, \theta)} \mu'(X, Z, \beta) \quad (2.2.4)
$$

Then the conditionally unbiased estimating equation for $\beta$ is the sum of (2.2.4) over the observed values.

$$
\sum_{i=1}^{n} \gamma(y_i, w_i, z_i, \beta \mid \sigma, \theta) \mu'(x_i, z_i, \beta) = 0 \quad (2.2.5)
$$

Here

$$
\gamma(y_i, w_i, z_i, \beta \mid \sigma, \theta) = \frac{Y - \mu(X, Z, \beta)}{\sigma^2 v^2(X, Z, \beta, \theta)}
$$

There are many methods to estimate the nuisance parameters $\sigma$ and $\theta$. Most of these methods are based on true replicates or on functions of squared residuals. One of the methods used in variance component modeling is to use the following estimating equation:
$$\sum \left\{ Y(y_i, w_i, z_i, \beta | \sigma, \theta) - \frac{n-k}{n} d(\hat{x}_i, z_i, \beta, \sigma, \theta) \right\}$$

Here \( d(\cdot) \) is the column vector of derivatives of \( \log(\sigma v(\hat{x}_i, z_i, \beta, \theta)) \) with respect to \((\sigma, \theta)\). For simplicity let \( \beta = (\beta^*, \beta_2) \), where \( \beta^* \) is the regression coefficient of the error-prone predictor variables, \( X \), and \( \beta_2 \) is the regression coefficient of the error-free predictors, \( Z \). Suppose that all nuisance parameters \( \beta_2, \sigma \) and \( \theta \) are already obtained, say \( \hat{\beta}_2, \hat{\sigma} \) and \( \hat{\theta} \), via these estimating equations (2.2.1a) and (2.2.1b), we have

$$M_i (\beta^*) = Y(y_i, w_i, z_i, (\beta^*, \hat{\beta}_2) | \hat{\sigma}, \hat{\theta}) \mu'_{sub}(\hat{x}_i, z_i, (\beta^*, \hat{\beta}_2))$$

Hence \( \mu'_{sub}(\hat{x}_i, z_i, (\beta^*, \hat{\beta}_2)) \) is the column vector of derivatives of \( \mu \) with respect to \( \beta^* \). There are two specific types of covariance estimates based on whether or not the variance model has been approximately correctly specified. For a sample of size \( n \), \( \hat{\beta} \) is asymptotically normally distributed with mean \( \beta \) and covariance matrix \( \frac{1}{n} D^{-1}_n C_n D^{-1}_n \), where,

$$D_n = \frac{1}{n} \sum_{i=1}^{n} \mu'(X_i, Z_i, \beta) \mu'(X_i, Z_i, \beta)' \left\{ \sigma^2 v^2(X_i, Z_i, \beta, \theta) \right\}^{-1}$$

$$C_n = \frac{1}{n} \sum_{i=1}^{n} \mu'(X_i, Z_i, \beta) \mu'(X_i, Z_i, \beta)' \frac{E[Y - \mu(X_i, Z_i, \beta)]^2}{\sigma^2 v^4(X_i, Z_i, \beta, \theta)}.$$ 

Here

$$\mu(x_i, z_i, \beta) = E(y_i | x_i, z_i).$$

If the variance model is correct, then \( E[Y_i - \mu(X_i, Z_i, \beta)]^2 = \sigma^2 v^2(X_i, Z_i, \beta, \theta) \) and \( D_n = C_n \).

The asymptotically correct covariance matrix is \( \frac{1}{n} \hat{D}_n^{-1} \) where \( \hat{D}_n^{-1} \) is defined as
\[
\hat{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mu'(X_i, Z_i, \beta) \mu'(X_i, Z_i, \beta) \left( \hat{\sigma}^2 V^2 \left( X_i, Z_i, \beta, \hat{\theta} \right) \right)^{-1}
\]

Sandwich method is another approach to estimate \( E \{ Y_i - \mu(X_i, Z_i, \beta) \}^2 \). In this case the estimated covariance matrix becomes \( \hat{D}_n^{-1} \hat{C}_n \hat{D}_n^{-1} \), where

\[
\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \mu'(X_i, Z_i, \beta) \mu'(X_i, Z_i, \beta) \frac{E\{ Y - \mu(X_i, Z_i, \beta) \}^2}{\hat{\sigma}^2 V^4 \left( X_i, Z_i, \beta, \hat{\theta} \right)}.
\]

In some situations, likelihood type extensions of sandwich approach is used to estimate \( E \{ Y_i - \mu(X_i, Z_i, \beta) \}^2 \); see Huber (1967), Schrader and Hettmansperger (1980), Kent (1982), Ronchetti (1982), and Li and McCullagh (1994). It is an extension of the theory of estimating equations, where estimating equation is assumed to correspond to a criterion function; that solves the estimating equation by minimizing the criterion function.

The choice among these approaches is the variance function when the variance model fits the data fairly well and the estimated covariance matrix is \( \frac{1}{n} \hat{D}_n^{-1} \). The estimators in this approach carry much fewer variables than the sandwich estimator. If the variance function seems inadequate for some reason, one might try to find a better one and then obtain the estimated covariance matrix with the better-fitted model. One can use the ordinary or weighted residual plots to understand the departures from the assumed mean and can use the absolute residual plot to see the deviation from the variance function.

2.3 Idea of Empirical Likelihood

Likelihood-based methods are of fundamental importance in the statistical inference. The maximum likelihood estimator (MLE) automatically provides an estimator that is asymptotically...
efficient in several senses when the model is correctly specified. Likelihood based methods are very effectively used to find efficient estimators and construct tests with good power properties. This approach is very flexible to deal with incompletely observed data, or distorted data or, data sampled with bias. This method also incorporates the information arising from outside of the data. But it is widely recognized that the validity of the likelihood approach depends on the assumption of the parametric form for the data distribution. The problem with parametric approach is that one might not know which parametric family to use and there is no reason to assume that a newly encountered set of data fits in a well-studied parametric family. This misspecification may cause likelihood estimates inefficient and make the test unsuccessful completely. This fact has spurred the development of the nonparametric method. One of the earliest ideas of treating the data distribution nonparametrically in statistical estimation and testing is to use empirical distribution of the data by comparing it with the family of distribution(s) implied by a statistical model. Empirical likelihood method is a nonparametric method introduced by Owen (1990) for the statistical inference. The empirical distribution function based on a sample is renowned to be the maximum likelihood estimate of the distribution from which the sample came. Owen (1990) proved that the empirical likelihood ratio statistic has an asymptotic \( \chi^2 \) - distribution, and therefore is useful for the interval estimation and hypothesis testing. Qin and Lawless (1994) discovered that the empirical likelihood method is also a powerful tool for the point estimation when the side information can be incorporated into constrained maximization of the empirical likelihood function. The first application of the concept behind the empirical likelihood was suggested by Hartley and Rao (1968) for a finite population and the formal application of empirical likelihood method in survey sampling was introduced by Chen and Qin (1993) under SRS. The pseudo empirical likelihood approach under
a general sampling design was discussed by Chen and Sitter (1999). This method allows the researchers to use the likelihood method without having the assumption of the data distribution. According to Owen (1990), this approach is an alternative likelihood type bootstrap method. However, Hall and Scala (1990) argued that the empirical likelihood is a serious competitor with contemporary methods, like several versions of the bootstrap such as percentile-t, accelerated bias correction, jackknife, infinitesimal jackknife etc. Each approach has its own advantages. According to Owen (1990), the advantages of empirical likelihood take place because it combines the reliability of the nonparametric approach with flexibility and effectiveness of the likelihood approach.

Example

To see the necessity of the empirical likelihood over the parametric likelihood and other nonparametric approaches, Owen (2001) considered the earthworm example. The earthworm, *Lumbricus herculeus* occurs abundantly in the region in Ann Arbor. This type of earthworm is widely used for laboratory teaching and biological investigation. In this study, Pearl and Fuller (2006) collected data from adult earthworms with fully developed clitellum. The worms were collected in rainy nights, in gardens and lawns near the campus of the University of Michigan. It was easy to pick up a large number of earthworms in a short time in this type of environment. The worms were brought to the laboratory, stupefied in alcohol to relax their muscles and straighten them to their full length. They are hardened in this straightened position. The following information was collected from each worm.

1. Total number of segments, known as somites in the body
2. Length of the body in centimeter
3. The number of somites from the anterior end of the worm to the most anterior somite of the clitellum

4. The number of somites included in the clitellum

5. The position of the genital openings

All worms with spiral segments were excluded from this research. Considerable care was taken to measure “Length of the body” to get reliable measurement of the length of an earthworm.

Table 2.1 shows the frequency distribution of the total number of somites in the body.
Table 2.1: Total Somites

<table>
<thead>
<tr>
<th>Number of Somites</th>
<th>Frequency</th>
<th>Number of Somites</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>1</td>
<td>130</td>
<td>1</td>
</tr>
<tr>
<td>80-88</td>
<td></td>
<td>131</td>
<td>4</td>
</tr>
<tr>
<td>89</td>
<td>2</td>
<td>132</td>
<td>7</td>
</tr>
<tr>
<td>90-92</td>
<td></td>
<td>133</td>
<td>4</td>
</tr>
<tr>
<td>93</td>
<td>1</td>
<td>134</td>
<td>3</td>
</tr>
<tr>
<td>94</td>
<td></td>
<td>135</td>
<td>5</td>
</tr>
<tr>
<td>95</td>
<td>1</td>
<td>136</td>
<td>6</td>
</tr>
<tr>
<td>96-98</td>
<td></td>
<td>137</td>
<td>9</td>
</tr>
<tr>
<td>99</td>
<td>2</td>
<td>138</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>139</td>
<td>6</td>
</tr>
<tr>
<td>101</td>
<td></td>
<td>140</td>
<td>5</td>
</tr>
<tr>
<td>102</td>
<td>2</td>
<td>141</td>
<td>17</td>
</tr>
<tr>
<td>103</td>
<td></td>
<td>142</td>
<td>27</td>
</tr>
<tr>
<td>104</td>
<td>3</td>
<td>143</td>
<td>21</td>
</tr>
<tr>
<td>105</td>
<td>1</td>
<td>144</td>
<td>24</td>
</tr>
<tr>
<td>106</td>
<td>1</td>
<td>145</td>
<td>37</td>
</tr>
<tr>
<td>107</td>
<td>2</td>
<td>146</td>
<td>32</td>
</tr>
<tr>
<td>108-112</td>
<td></td>
<td>147</td>
<td>31</td>
</tr>
<tr>
<td>113</td>
<td>1</td>
<td>148</td>
<td>29</td>
</tr>
<tr>
<td>114</td>
<td>3</td>
<td>149</td>
<td>26</td>
</tr>
<tr>
<td>115</td>
<td>3</td>
<td>150</td>
<td>51</td>
</tr>
<tr>
<td>116</td>
<td></td>
<td>151</td>
<td>15</td>
</tr>
<tr>
<td>117</td>
<td></td>
<td>152</td>
<td>15</td>
</tr>
<tr>
<td>118</td>
<td>2</td>
<td>153</td>
<td>12</td>
</tr>
<tr>
<td>119</td>
<td></td>
<td>154</td>
<td>9</td>
</tr>
<tr>
<td>120</td>
<td>5</td>
<td>155</td>
<td>9</td>
</tr>
<tr>
<td>121</td>
<td></td>
<td>156</td>
<td>10</td>
</tr>
<tr>
<td>122</td>
<td>1</td>
<td>157</td>
<td>4</td>
</tr>
<tr>
<td>123</td>
<td>2</td>
<td>158</td>
<td>-</td>
</tr>
<tr>
<td>124</td>
<td>2</td>
<td>159</td>
<td>2</td>
</tr>
<tr>
<td>125</td>
<td>5</td>
<td>160-161</td>
<td>-</td>
</tr>
<tr>
<td>126</td>
<td>2</td>
<td>162</td>
<td>2</td>
</tr>
<tr>
<td>127</td>
<td>1</td>
<td>163</td>
<td>-</td>
</tr>
<tr>
<td>128</td>
<td>5</td>
<td>164</td>
<td>1</td>
</tr>
<tr>
<td>129</td>
<td>4</td>
<td>Total</td>
<td>487</td>
</tr>
</tbody>
</table>
The distribution is shown graphically in Figure 2.1

![Histogram of Worms vs Somites](image)

**Figure 2.1 Histogram of the Number of Somites of 487 Earthworms Reported in (1905)**

From the histogram, it is virtually apparent that the earthworms with many fewer segments than the average are more common than worms with many more segments. The distribution of somites is skewed and has a heavier tail to the left than it has to the right. There are mildly skewed, very skewed and extremely skewed distributions, and the coefficient of skewness for random variable $X$, is defined as,

$$
\gamma = \frac{E\left[(X - E(X))^3\right]}{E\left[(X - E(X))^2\right]^{3/2}}
$$
for a symmetric distribution $\gamma = 0$, and for the earthworm-somites data the estimated skewness is $\hat{\gamma} = -2.18$. The estimated $\hat{\gamma}$ is the skewness of the empirical distribution, which places probability $\frac{1}{487}$ on each of the 487 observed values. The kurtosis is defined as,

$$\kappa = \frac{E[(X - E(X))^4]}{E[(X - E(X))^2]^2} - 3$$

The kurtosis is zero for a normally distributed data and positive value of the kurtosis defines a distribution with heavier tails than the normal distribution and negative value describes a lighter tail than normal. The estimated kurtosis for the earthworm data is $\kappa = 5.86$.

The most popular parametric methods begins by assuming the normal distribution $N(\mu, \sigma^2)$ for the data where $\gamma = \kappa = 0$; and it cannot be used to draw inference on skewness and kurtosis in earthworm data.

If the data is not normally distributed then the method based on normal distribution gives asymptotically reliable inference for the mean. In this case the data must have a finite variance.

The true variance of the sample mean is $\sigma^2/n$, the estimated value of $\sigma^2$ approaches the true value and the confidence intervals based on this estimated variance are asymptotically reliable.

The maximum likelihood estimate of the variance is, $\hat{\sigma}^2 = \frac{S^2}{n}$ and $S = \sum_{i=1}^{n} (X_i - \bar{X})^2$,

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$; $\hat{\sigma}^2$ converges to $\sigma^2$ as $n \to \infty$ and $Var(s^2) = \sigma^4 \left( \frac{2}{n-1} + \frac{\kappa}{n} \right)$

where $s^2 = \frac{S}{n-1}$. The normal distribution forces $\kappa = 0$. If the sampling distribution has $\kappa \neq 0$, the normal theory cannot be used as an asymptotically correct estimate of $n \cdot Var(s^2)$ or
and the resulting confidence intervals for \( \sigma^2 \) do not approach to nominal coverage levels as \( n \) increases.

If a normal distribution is inadequate, perhaps some other distribution is better. It is really difficult to choose a parametric family that fit the data and allow both skewness and kurtosis move freely. But it would be still more difficult to find a parametric family in which inferences for the skewness and kurtosis would be reliable, even if the data did not come from a member of that parametric family. If we replace the parametric method with a nonparametric one, like the empirical likelihood method, improved estimates comes at a cost of reduced power. The use on nonparametric method is in the line with Tukey’s quote “It is better to be approximately right, than exactly wrong”. For this type of situation, we will consider the empirical likelihood function since empirical likelihood tests have especially good power properties, and it provides data determined shapes for the confidence regions, that is, the shape of a resulting confidence region is determined automatically by the data. It easily incorporates known constrains on parameters and adjusts for biased sampling schemes. Use of this method makes it easier to combine data from multiple sources, sometimes from different sampling schemes. The only disadvantage of the empirical likelihood is that sometimes it is computationally challenging to optimize likelihood.

2.4 Empirical Likelihood Inference for a Scalar Random Variable

For a random variable \( X \in IR \), the cumulative distribution function (CDF) is defined as,

\[
F(x) = \Pr(X \leq x), \ -\infty < x < \infty
\]

Define \( F(x-) = \Pr(X < x) \); we have,

\[
\Pr(X = x) = F(x) - F(x-)
\]

Again, consider the indicator function,
\[ 1_{A(x)} = \begin{cases} 1, & \text{if } A(x) \text{ is true} \\ 0, & \text{otherwise} \end{cases} \]

Then, the empirical cumulative distribution function (ECDF) of \( X_1, X_2, \ldots, X_n \) becomes,

\[ F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x}, -\infty < x < \infty \]

Therefore, the nonparametric likelihood of \( F \) is,

\[ L(F) = \prod_{j=1}^{n} \left[ F(X_j) - F(X_j^-) \right] \]

Here, it is assumed that \( X_1, X_2, \ldots, X_n \in IR \) are independent with common CDF \( F_0 \).

Hence \( L(F) \) is the probability of getting exactly the observed sample values \( X_1, X_2, \ldots, X_n \) from the CDF \( F \).

**Theorem 2.1**

Let \( X_1, X_2, \ldots, X_n \in IR \) be independent random variables with a common CDF \( F_0 \). Let \( F_n \) be their ECDF and \( F \) be any CDF. If \( F \neq F_n \), then \( L(F) < L(F_n) \).

**Proof:**

Let \( z_1 < z_2 < \ldots < z_m \) be the distinct values in \( \{X_1, X_2, \ldots, X_n\} \) and let \( n_j \geq 1 \) be the number of \( X_i \) that are equal to \( z_j \).

Let \( p_j = F(z_j) - F(z_j^-) \) and \( \hat{p}_j = \frac{n_j}{n} \).

If \( p_j = 0 \) for any \( j = 1, 2, \ldots, m \), then \( L(F) = 0 < L(F_n) \). Therefore, we consider that all \( p_j > 0 \) and for at least one \( j \), \( p_j = \hat{p}_j \). We know \( \log(x) \leq x - 1 \) for all \( x > 0 \) and equality only when \( x = 1 \). Therefore,
\[
\log \left( \frac{L(F)}{L(F_n)} \right) = \sum_{j=1}^{m} n_j \log \left( \frac{p_j}{\hat{p}_j} \right)
\]
\[
= n \sum \hat{p}_j \log \left( \frac{p_j}{\hat{p}_j} \right)
\]
\[
< n \sum_{j=1}^{m} \hat{p}_j \left( \frac{p_j}{\hat{p}_j} - 1 \right) \leq 0
\]

That is,
\[
\log \left( \frac{L(F)}{L(F_n)} \right) \leq 0;
\]

Therefore, if \( F \neq F_n \), \( L(F) < L(F_n) \).

\subsection*{2.4.1 Nonparametric Likelihood Ratio}

In the parametric approach, the inference on hypothesis tests and confidence regions depends on the likelihood ratio. We reject the hypothesis \( \gamma_0 = \gamma \) and exclude \( \gamma \) from the confidence region for \( \gamma_0 \) when \( L(\gamma) \) is much smaller than \( L(\hat{\gamma}) \). Wilk’s theorem provides that
\[-2 \log \left( \frac{L(\gamma_0)}{L(\hat{\gamma})} \right)\]
approaches to a chi-square distribution as \( n \to \infty \) under mild regularity condition; where the degrees of freedom in the chi-square distribution are equal to the dimension of \( \gamma \). For a distribution \( F \), the nonparametric likelihood ratio for hypothesis tests and confidence intervals is defined as,
\[R(F) = \frac{L(F)}{L(F_n)}\]

Suppose, we are interested in \( F \) through \( \theta = G(F) \), where \( G(\cdot) \) is a function of distributions and \( F \) is a member of a set of distributions \( F \). Let \( \theta_0 = G(F_0) \). Then the nonparametric maximum likelihood estimate of \( \theta \) is \( \hat{\theta}_0 = G(F_n) \) and the profile likelihood ratio function is defined as,
\[ \Re(\theta) = \sup \left\{ R(F) \mid G(F) = \theta, F \in \mathcal{F} \right\} \]  

(2.4.1)

Empirical likelihood hypothesis tests reject the hypothesis \( G(F_0) = \theta_0 \), when \( \Re(\theta_0) < \tau_0 \), for some threshold value \( \tau_0 \). The empirical likelihood confidence region becomes,

\[ \{ \theta \mid \Re(\theta) \geq \tau_0 \} \].

2.4.2 Tied Observations

If there is no tie in the data set, then the empirical likelihood ratio becomes,

\[ R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} p_i \]

(2.4.2a)

where the distribution \( F \) places probability \( p_i \geq 0 \) on the value \( X_i \in \mathcal{F} \) and \( \sum_{i=1}^{n} p_i \leq 1 \) and empirical likelihood function \( L(F) = \prod_{i=1}^{n} p_i \).

Suppose there are ties in the data set and \( X_j \) arises \( n_j \geq 1 \) times in the sample with the probability \( p_j \) under \( F \). Let \( k \) be the number of distinct values in the data, then the empirical likelihood ratio becomes

\[ R(F) = \frac{L(F)}{L(F_n)} = \prod_{j=1}^{k} \left( \frac{p_j}{\hat{p}_j} \right)^{n_j} = \prod_{j=1}^{k} \left( \frac{np_j}{n_j} \right)^{n_j} \]

(2.4.2b)

2.4.3 Empirical Likelihood (EL) for a Univariate Mean

Theorem 2.2

Let \( X_1, X_2, \ldots, X_n \) be independent random variables with a common CDF \( F_0 \). Let \( \mu_0 = E(X_i) \) and suppose \( 0 < Var(X_i) < \infty \). Then \( -2 \log(\Re(\mu_0)) \) converges in distribution to \( \chi^2_0 \) as \( n \to \infty \).
This theorem provides an asymptotic result for tests that reject the value $\mu_0$ at the $\alpha$-level, when $-2 \log(\mathcal{R}(\mu_0)) > X^2_{(1)}(1-\alpha)$. The unrejected value of $\mu_0$ form a $100(1-\alpha)$% confidence region. The EL confidence region for a mean is always an interval.

### 2.4.4 Coverage Accuracy

A $100(1-\alpha)$% EL confidence interval is, $\{\mu \mid -2 \log(\mathcal{R}(\mu)) \leq X^2_{(1)}(1-\alpha)\}$.

That is equivalent to

$$\left\{ \mu \left| \mathcal{R}(\mu) \leq \exp\left(\frac{X^2_{(1)}(1-\alpha)}{2}\right) \right. \right\}$$

As $n \to \infty$,

$$\Pr\left[ -2 \log(\mathcal{R}(\mu)) \leq X^2_{(1)}(1-\alpha) \right] - (1-\alpha) \to 0$$

The rate $\left(\frac{1}{n}\right)$ of convergence is the same as the confidence interval based on the parametric likelihood, the jackknife, and the simpler bootstrap methods.

### 2.4.5 Power and Efficiency

One can assess the power of EL through the curvature of $\mathcal{R}$ at the nonparametric maximum likelihood estimate, $X$. For the large sample sizes, $\log(\mathcal{R}(\mu)) = -n \frac{\sigma^2_0}{2} \left(\mu - \overline{X}\right)^2$ for $\mu$ near $\overline{X}$, where $\sigma^2_0 = Var(X)$. The greater the curvature in the quadratic, the shorter the confidence intervals for a given level of coverage, and hence the greater power. It can be shown that,

$$-2 \log\left(\mathcal{R} \left( \mu_0 + \frac{\tau \sigma_0}{\sqrt{n}} \right) \right) \to X^2_{(1)}(\tau^2)$$

It is convergence in distribution, where $\tau^2$ is a noncentrality parameter. At $\mu \neq \mu_0$, the nonparametric inference will have powers roughly same as the parametric inferences.
2.4.6 Computational Chores for a Univariate Mean

The empirical likelihood inference for a univariate mean requires the following computational steps: to test whether \( \mu = \mu_0 \), one needs to compute \( \mathcal{R}(\mu_0) \), find two values of \( \mu \) that solve \( \mathcal{R}(\mu) = \tau_0 \), where \( \tau_0 \) is a threshold value to set the confidence limit for \( \mu \). Before computing \( \mathcal{R}(\mu) \), order the values \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \). Eliminate the trivial cases, for example, if \( \mu < X_{(1)} \) or \( \mu > X_{(n)} \) there is no weight \( w_i \geq 0 \) summing to 1 for which

\[
\sum_{i=1}^{n} w_i X_i = \mu.
\]

In this situation, \( \log(\mathcal{R}(\mu)) \) becomes either \(-\infty\) or 0. Again if \( \mu = X_{(1)} < X_{(n)} \), or \( \mu = X_{(n)} > X_{(1)} \), then \( \mathcal{R}(\mu) = 0 \) and if \( X_{(1)} = X_{(n)} = \mu \), then \( \mathcal{R}(\mu) = 1 \).

Consider only the nontrivial case, where \( X_{(1)} < \mu < X_{(n)} \). We need to maximize likelihood ratio,

\[
\mathcal{R}(\mu) = \prod_{i=1}^{n} n w_i \quad (2.4.3a)
\]

where \( w_i \geq 0 \), \( \sum_{i=1}^{n} w_i \leq 1 \) and the distribution function \( F \) puts the probability, \( \sum_{j: X_j = x_j} w_j \) on \( X_j \), or equivalently maximize the objective function,

\[
\mathcal{R}(\mu) = \sum_{i=1}^{n} \log(nw_i) \quad (2.4.3b)
\]

over \( w_i \geq 0 \) subject to the condition \( \sum_{i=1}^{n} w_i = 1 \) and \( \sum_{i=1}^{n} w_i X_i = \mu \). Since the objective function is a strictly concave function on a convex set of weight vectors, a global maximum exists. Again maximum does not have any \( w_i = 0 \), so it is an interior point of the domain.

Consider \( \lambda \) and \( \delta \) be the Lagrange multipliers and the score function becomes

\[
H = \sum_{i=1}^{n} \log(nw_i) - n \lambda \sum_{i=1}^{n} w_i (X_i - \mu) + \delta \left( \sum_{i=1}^{n} w_i - 1 \right) \quad (2.4.4)
\]
After taking the partial derivatives with respect to $w_i$’s and set it equal to zero, we obtain $\delta = -n$

Therefore,

$$w_i = \frac{1}{n} \frac{1}{1 + \lambda(X_i, \mu)}$$

Furthermore, the value of $\lambda$ may be found by numerical search, and $\lambda = \lambda(\mu)$ solves the following equation,

$$\frac{1}{n} \sum \frac{x_i - \mu}{1 + \lambda(X_i - \mu)} = 0$$

We have to set an interval that contains $\lambda(\mu)$ before the numerical search using every $w_i > 0$ and $w_i < 1$, this interval becomes,

$$\frac{1 - \frac{1}{n}}{\mu - X_{(n)}} < \lambda(\mu) < \frac{1 - \frac{1}{n}}{\mu - X_{(1)}}$$

Then we set an algorithm that successively refines the interval for $\lambda$ until the endpoints agree to a specified tolerance. To set a confidence interval for $\mu$, one needs to find the upper and lower limits, $\mu_+$ and $\mu_-$ for which $\Re(\mu_+, \mu_-) = \tau_0 \in (0,1)$, that is, $X_{(1)} \leq \mu_- \leq \bar{X} \leq \mu_+ \leq X_{(n)}$. Then search for the $\mu$ so that $\Re(\mu) = \tau_0$.

2.5 Empirical Likelihood for Random Vectors

Let $X_1, X_2, \ldots, X_n \in IR^p$ be independent random vectors for some $p \geq 1$, with a common distribution function $F_0$. In this case, we describe distributions by the probabilities that they attach to each set. Therefore, $F(A) = \Pr(X \in A)$ for $X \sim F$ and $A \subseteq IR^p$, or, in other words, $F(\{X_i\})$ is the probability of getting the value $X_i$ in a sample from $F$, $x_i$ is the observed value of $X_i$. Define $\zeta_x(A) = 1_{x \in A}$. The empirical likelihood function of $X_1, X_2, \ldots, X_n$ is,
\[ F_n = \frac{1}{n} \sum_{i=1}^{n} \hat{F}(X_i) \]

The nonparametric likelihood of the distribution function \( F \) is

\[ L(F) = \prod_{i=1}^{n} F(X_i) \]

And the likelihood function becomes

\[ R(F) = \frac{L(F)}{L(F_n)} \]

It can be used to construct nonparametric confidence regions and tests. The function \( R(F) \) has similar properties like parametric likelihood ratio functions.

2.5.1 Empirical Likelihood for a Multivariate Mean

The profile likelihood ratio function on \( IR^p \) is defined as,

\[ \mathcal{R}(\mu) = \max \left\{ \prod_{i=1}^{n} nw_i \mid \sum_{i=1}^{n} w_i X_i = \mu, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\} \]

The confidence interval for multivariate mean is a subset of \( IR^p \) which is,

\[ C_{\tau,n} = \left\{ \sum_{i=1}^{n} w_i X_i \mid \prod_{i=1}^{n} m w_i \geq \tau, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\} \]
Suppose our interest is in the regression of a response variable $Y$, on covariates $X$ which is measured with errors. Their true values are unobservable or latent and can be observed only through an imprecise measurement or surrogate variable $W$, whereas the other covariates $Z$ are perfectly measured or observed. Regression with measurement error refers to the problems that arise from using $W$ instead of $X$ in the model of interest. If one variable is measured with error, point and interval estimates of its effect are often, although not always, underestimated. However, if two or more variables are included in a regression model and at least one of them is measured with error, the point and interval estimates of any of the effects may be underestimated or overestimated, including even those estimates corresponding to variables measured without error. In general, ignoring the measurement error results in biased estimates of the parameters of the regression of $Y$ on $X$. Once data from error prone studies are available, investigators have all the information necessary to correct their biased point and interval estimates of the parameters caused by measurement error in one or more covariates. Investigators have proposed a variety of methods to correct or reduce the bias due to measurement error in one or more continuous covariates. This problem was noted and studied as early as 1877 by Adcock. The problem of covariate measurement error has been reviewed in Fuller (1987) for the linear regression and Carroll, Ruppert and Stefanski (1995) for the nonlinear regression. It occurs in many applications (Carroll, Spiegelman, Lan, Bailey and Abbott (1984), Rosner, Willett and Spiegelman (1989), Pierce, Stram, Vaeth and Schafer (1992), Prentice (1996) and Gustafson (2004)) from a Bayesian perspective. Assessing the impact and possible corrections of the measurement error requires an understanding of the measurement process. This involves the formulation of a conceptual model for the relationship between the true variable $X$ and its surrogate $W$. Classical error structure
arises when \( W = X + U \), and \( U \) is independent of \( X \). This structure is appropriate for an imprecise measurement device that adds noninformative noise to the true value of \( X \). Berkson error structure arises when \( X = W + U \) and \( U \) is independent of \( W \) (Berkson 1950). This structure was initially proposed for controlled experiments, in which a nominal level of a treatment was prescribed to an experimental unit, but the actual level applied was the nominal level plus some noninformative noise. In the simple linear model, classical error results in an attenuation of the regression slope, while Berkson error allows for unbiased estimation of the regression parameters (see, for example, Madansky 1959, Cochran 1965). In more complex cases, there is no simple pattern and both structures can result in attenuation, inflation, or can even induce a curvature in the regression of \( Y \) on \( W \) (Fuller 1987, Reeves et al. 1998).

Investigators have proposed an intuitive, noniterative method which is called regression calibration, to correct point and interval estimates due to measurement error in one or more covariates. The regression-calibration method was suggested by Carroll and Stefanski (1990) and also by Glesjer (1990). This method is also discussed in terms of proportional hazards models by Prentice (1982) and in terms of generalized linear models by Armstrong (1985). Carroll, Ruppert, and Stefanski (1995) discussed the regression-calibration method for correcting bias in fitting generalized linear measurement error models. The basis of the regression calibration algorithm for measurement error analysis is the construction of the calibration model for generation of estimated covariate values for the unknown covariates.

### 3.1 Overview of Regression Calibration

Regression calibration seeks to estimate regression models with measurement error in explanatory variables. The misspecified predictor variable is replaced by its conditional expectation, given a surrogate variable, in an estimation procedure that would have been used if
the true values were available. Because of its transparency, ease of use, and apparently good operational characteristics, the regression calibration has emerged as an important tool for estimation of regression parameters in the presence of measurement errors in explanatory variables.

Let $Y$ be the response variable, $X$ be the predictor variables measured with error, $W$ be a measurement or surrogate for $X$, and $Z$ be additional explanatory variables free of measurement error. The regression calibration estimator, in its most transparent form, uses the regression estimator that would have been used if $X$ were known exactly (for linear, generalized linear, or failure time regression models), but with the missing $X$ replaced by $E\left( X \mid Z, W \right)$ (Carroll et al. 1995, Ch. 3). Regression calibration can be widely used in many disciplines. Two important areas of application are nutritional epidemiology (Willett et al. 1992, Binham et al. 2003) and radiation health epidemiology (Pierce et al. 1990, Stram et al. 1999, Schafer 2001).

The previous one primarily involves logistic and failure time regression models for binary health responses on diet and nutrition predictor variables, which are measured imprecisely. The second one involves failure time regression models that are linear or quadratic functions of dose of radiation, which is observed through an imprecise estimate.

It is easy to notice the justification for the regression calibration in the simple linear regression. If the regression of $Y$ on $X$ is linear,

$$E( Y \mid X ) = \beta_0 + \beta_1 X,$$

then

$$E( Y \mid W ) = E\{ E( Y \mid X ) \mid W \} = \beta_0 + \beta_1 E( X \mid W ).$$

Since the coefficients in the regression of $Y$ on $E( X \mid W )$ are the same as those in the regression of interest $E( Y \mid X )$. The practical attention can be switched to the regression of $Y$ on $E( X \mid W )$.
This also shows that the naïve regression of $Y$ on $W$ will lead to the biased estimation of the regression coefficients if $W$ is not the same as $E(X \mid W)$, as is the case under the classical measurement error model. In the classical model, $W$ is the sum of $X$ and a random measurement error that is independent of $X$, and the estimated slope of the regression line is biased towards zero (see, for example, Madansky 1959, Cochran 1965). On the other hand, under the Berkson error model, in which $E(X \mid W)=W$ follows from the model definition, the usual estimators of regression coefficients are unbiased (Berkson 1950).

If $E(Y \mid X,Z)$ is not linear in $X$, the simple substitution of $E(X \mid W,Z)$ in place of $X$ in the regression model $E(X \mid W,Z)$ leads to an approximate model for $E(X \mid W,Z)$. Estimation based on this substitution is, in general, biased and inconsistent (Carroll et al. 1995). However, under additional assumptions that depend on the particular model, it is approximately consistent, and the approximation may be improved with a second order approximation to $E(Y \mid X,Z)$ at $X = E(X \mid W,Z)$ (Carroll and Stefanski 1990, Kuha 1994). For generalized linear models, $Var(Y \mid W)$ will not have the same form as $Var(Y \mid X)$ and some attention to proper “weighting” is necessary for efficient estimation.

Uncertainty in $E(X \mid W,Z)$ is not negligible in most epidemiological studies. For example, in diet and health research from the Nurse’s Health Study, a primary data set consists of almost 90,000 nurses. Investigators consider the regression of health outcomes on explanatory variables, $X$ associated with diet, such as total fat intake. Surrogate, $W$, for this type of variable is measured on all nurses in the primary study, and $E(X \mid W,Z)$ is taken to be an estimated mean from a regression model fit to a calibration dataset of 173 nurses (Willett et al. 1992). The estimated correlation between $X$ and $W$ is low – typically, between 0.4 and 0.6. The values
used as $E(X \mid W, Z)$ for each of the nurses in the primary data set, therefore contain a component of uncertainty due to the sampling error from the regression estimation on the calibration set.

### 3.1.1 The Regression Calibration Algorithm

Regression calibration inference tries to find estimate of regression models with measurement error in predictor variables by replacing the error-prone variable by its conditional expectation, given a surrogate variable, in an estimation procedure that would have been used if the true variables were available. The regression calibration method is a simple approach wherein one needs only to develop and fit the calibration model for the regression of the unknown covariates $X$ on $(Z, W)$. This is accomplished using replication, validation, or instrumental data in place of the unknown $X$. This first-stage regression results in a calibration function for estimating $X$. The unobserved covariates are then replaced by their predicted values from the calibration model in a standard analysis. Finally, the standard errors are adjusted to account for the estimation of the unknown covariates. The typical approach is to calculate standard errors using bootstrap or sandwich methods. The algorithm of regression calibration is as follows; see Carroll et al. 2006:

- Estimate the regression of $X$ on $(Z, W)$, $E(X \mid W, Z, \gamma)$, depending on parameters $\gamma$, which are estimated by $\hat{\gamma}$.

- Replace the unobserved $X$ by its estimate $E(X \mid W, Z, \hat{\gamma})$, and then run a standard analysis to obtain parameter estimates.

- Adjust the resulting standard errors to account for the estimation of $\gamma$, using either bootstrap or sandwich method.
It should be noted that the regression calibration model is a working model for the observed data, and in not the same as the actual mean for the observed data, but in many situation, is only modestly different.

3.1.2 The Regression Calibration Estimator for Linear Regression

Suppose $Y$ is the response variable, $X$ is the error prone predictor variable, $W$ is the surrogate for the explanatory variable, $(\beta_0, \beta_1, \alpha_0, \alpha_1)$ are unknown regression coefficients, and $(\varepsilon, U)$ are random variables. Then one can define,

\[
Y = \beta_0 + \beta_1 X + \varepsilon \quad (3.1.1)
\]

\[
X = \alpha_0 + \alpha_1 W + U \quad (3.1.2)
\]

Now replacing $X$ in (3.1.1) by (3.1.2),

\[
Y = (\beta_0 + \alpha_0 \beta_1) + \alpha_1 W + (\varepsilon + \beta_1 U) \quad (3.1.3)
\]

Rewrite (3.1.3) in the following simple form,

\[
Y = \delta_0 + \delta_1 W + u \quad (3.1.4)
\]

where, $\delta_0 = (\beta_0 + \alpha_0 \beta_1)$, $\delta_1 = \alpha_1 \beta_1$ and $u = (\varepsilon + \beta_1 U)$. Now (3.1.4) is the regression of $Y$ on $W$.

Suppose that there is a primary sample consisting of observations $(y_i, w_i)$, $i = 1, 2, \ldots, n$ and an independent calibration sample consisting of observations $(x_i, w_i)$ $i = n + 1, n + 2, \ldots, n + m$, $(\varepsilon, U)$ are independent random errors with means equal to 0, random variables associated with different values of $i$ are independent of one another; and the error structure is non-differential, meaning $f(Y | X, W) = f(Y | X)$. 
For this model, the regression calibration estimator of $\beta_i$ can be defined in the following two different but equivalent approaches. One approach consists of estimating $(\alpha_0, \alpha_i)$ by $(\hat{\alpha}_0, \hat{\alpha}_i)$ from the external calibration sample, calculating $\hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_i w_i$ for each observation in the primary sample, and estimating the slope of the regression of $y_i$ on $\hat{x}_i$ using least squares (Carroll et al. 1995, Chapter 3). Then, the regression calibration estimator takes the form;

$$
\hat{\beta}_{rc} = \frac{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2} = \hat{\delta}_i \hat{\alpha}_i
$$

where $\hat{\delta}_i$ and $\hat{\alpha}_i$ are the least squares estimators of the slope of $Y$ on $W$ based on the primary data, and of $X$ on $W$ based on the calibration data, respectively. The other approach arrives at the same estimator directly from equation (3.1.3), by solving $\delta_i = \alpha_i \beta_i$ for $\beta_i$ and substituting $\delta_i$ and $\alpha_i$ by their respective estimators (Rosner et al. 1989). More generally, as long as the regression of $X$ on $W$ is linear, both approaches yield $\hat{\beta}_{rc}$. This is true, for example, when the regression of $Y$ on $X$ is a generalized linear model.

It is not possible to derive the sampling distribution of the estimator $\hat{\beta}_{rc}$ without additional distributional assumptions. Asymptotic properties depend on the theory of estimating equations (Carroll et al. 1995, Appendix). In an asymptotic setting in which $n$ and $m$ both increase to infinity, the sampling distribution of the estimator $\hat{\beta}_{rc}$ converges to a normal distribution (Carroll and Stefanski 1990). The mean is $\beta_i$ and, under the additional assumption that the variances of $(e, U)$ are constants, the variance is,

$$
Var(\hat{\beta}_{rc}) = \frac{\hat{\delta}_i^2}{\hat{\alpha}_i^3} \sigma_\alpha^2 + \frac{1}{\hat{\alpha}_i} \sigma_\delta^2
$$

(3.1.6)
\[ \text{Var} \left( \hat{\delta}_1 \right) = \sigma_\delta^2 \quad \text{and} \quad \text{Var} \left( \hat{\alpha}_1 \right) = \sigma_\alpha^2. \]

Tests and confidence intervals are based on this asymptotic distribution, with unknown parameters replaced by their estimates and the bootstrap method can also be used (Carroll et al. 1995, Rosner et al. 1989). Confidence intervals based on asymptotic normality have been widely used and have been implemented in readily available software (Spiegelman et al. 1997). However, although asymptotically correct as \( m \to \infty \), they may not have very desirable finite sample properties. The actual sampling distribution of \( \hat{\beta}_{rc} \) can be very skewed, even for relatively large sample, while the interval is symmetric about \( \hat{\beta}_{rc} \).

A bootstrap confidence interval, based on the percentiles of the bootstrap replications, perform better in this situation. Besides, the expectation of the estimated variance given by (3.3.6) typically does not exist. Therefore, for any finite sample size, a confidence interval with non-zero coverage probability (e.g. a 95% CI) has an expected infinite length, which is a common feature of confidence interval in the measurement error problem (Gleser and Hwang 1987).

3.1.3 Example of Regression Calibration

The reason of the example (see, Fuller 2006) is to show an application of regression calibration to the regression model where predictor variable is measured with error. The data collected by the Iowa Conservation Commission. The Table 3.1 contains indexes of the number of hen pheasants in Iowa at two times during the year. These indexes are based on the average number of birds sighted by trained observers traveling a number of specific routes in late April and early May, and again in August. Both measures are subject to error for two reasons: first, the routes are a sample of all possible routes in Iowa. Second, observers cannot be expected to sight all pheasants along the route. In August, it is relatively easy to sight hen pheasants because most
pheasants are accompanied by a brood of chicks at that time. On the basis of other analyses, it has been estimated that the error variance for the spring count is about six times that for August. For this example it is assumed that the error variance ratio is known. It is not strictly expected that the number of pheasants in August is a perfect linear function of the number in the spring, but carrying such a specification is considered as a reasonable approximation. One should notice that the estimated error variance contains components due to variations in the survival rates as well as that due to observational and sampling errors.
Table 3.1 Index of the Number of Hen Pheasants in Iowa

<table>
<thead>
<tr>
<th>Year</th>
<th>August Hens, $Y_i$</th>
<th>Spring Hens, $X_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1975</td>
<td>6</td>
<td>6.6</td>
</tr>
<tr>
<td>1974</td>
<td>9.8</td>
<td>12.3</td>
</tr>
<tr>
<td>1973</td>
<td>10.8</td>
<td>11.9</td>
</tr>
<tr>
<td>1972</td>
<td>9.7</td>
<td>11.9</td>
</tr>
<tr>
<td>1971</td>
<td>9.3</td>
<td>12</td>
</tr>
<tr>
<td>1970</td>
<td>9.2</td>
<td>9.6</td>
</tr>
<tr>
<td>1969</td>
<td>6.9</td>
<td>7.5</td>
</tr>
<tr>
<td>1968</td>
<td>8.1</td>
<td>10.9</td>
</tr>
<tr>
<td>1967</td>
<td>8.7</td>
<td>10.4</td>
</tr>
<tr>
<td>1966</td>
<td>8.7</td>
<td>10.2</td>
</tr>
<tr>
<td>1965</td>
<td>7.4</td>
<td>7.4</td>
</tr>
<tr>
<td>1964</td>
<td>10.1</td>
<td>11</td>
</tr>
<tr>
<td>1963</td>
<td>10</td>
<td>11.8</td>
</tr>
<tr>
<td>1962</td>
<td>7.3</td>
<td>8.2</td>
</tr>
</tbody>
</table>

Source: Data from Iowa Conservation Commission. Data for 1962-1968 have been adjusted.
Figure 3.1 Scatter plot of Spring hens versus August hens
To analyze the data, the following model is considered;

$$\text{August Hens} = \beta_0 + \beta_1 \left( \text{Spring Hens} \right) + \varepsilon$$

Since the exact number of spring hens cannot be observed, the investigators observed $W_i$.

Assume that,

$$\text{Observed Spring Hens, } w_i = \text{Unobserved True, } x_i + u_i,$$

and

$$(x_i, \varepsilon_i, u_i) \sim \text{NI} \left( \mu_x, 0 \right), \text{diag} \left( \sigma_{xx}, \frac{1}{6} \sigma_{uu}, \sigma_{uu} \right)$$

The estimation by the regression calibration is done as follows.

Table 3.2 Comparison between the Naïve and Regression Calibration Estimators

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept, $\hat{\beta}_0$</th>
<th>Slope, $\hat{\beta}_1$</th>
<th>$SE(\hat{\beta}_0)$</th>
<th>$SE(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naive</strong></td>
<td>2.14227</td>
<td>0.64941</td>
<td>0.8451</td>
<td>0.08275</td>
</tr>
<tr>
<td><strong>Regression</strong></td>
<td>1.1158</td>
<td>0.75158</td>
<td>0.95929</td>
<td>0.0092478</td>
</tr>
</tbody>
</table>

It can be seen from the Table 3.2 ignoring the measurement error the nuisance estimate of $\beta_0$, substantially differs from that of obtained from corrected estimate. Again there is also a difference between the estimated slopes calculated from the naïve and regression calibration approach.

3.2 Overview of Simulation Extrapolation

A common problem which complicates the statistical analysis is the inability to directly measure or precisely measure some of the important predictor variables. In epidemiological studies, for example, a person’s blood pressure varies significantly over time, and thus a
particular reading is not true measurement of his or her long-term average blood pressure (Palca 1990 (a), 1990(b), McMahon et al. 1990). Similar problems take places in the determination of environmental risk factors, for example, the quality of particulate matter measurements is frequently cited in the debate about the health effects of airborne particulate substance (Lipfert and Wyzga 1995a, 1995b; Schwartz et al. 1996). The importance of these and similar problems has encourage considerable statistical research on methods for the analysis of data measured with error, many of which are described by Fuller (1987) or Carroll et al. (1995).

In the linear regression, as well as nonlinear models such as the generalized linear models, it is well known that the naive estimator that ignores the measurement error leads to inconsistent regression parameter estimates. Correction for this bias in linear regression has a long history, but the analysis of the nonlinear measurement error models is of more recent vintage. One of the more useful general methods is the regression calibration; it simplifies the usual correction for attenuation in linear regression, and consistent for the slopes in models with mean which is an exponential function of a linear combination of the predictors, and is very nearly consistent for slopes in most applications. In general, regression calibration estimates are easy to compute, have straightforward standard error estimates (Carroll, Stefanski 1990, 1995), and are amenable to the bootstrap analysis.

An alternative general method has been proposed by Cook and Stefanski (1994) called SIMEX that stands for the simulation and extrapolation. This idea depends on computer simulation to generate parameter estimates. Among many other methods, the SIMEX method has become a useful tool for correcting effect estimates in the presences of additive measurement error. The method is especially helpful for complex models with a simple measurement error structure. The same basic idea of the simulation and extrapolation is transferred to the case of
misclassification (MC–SIMEX) by Küchenhoff et al. (2007). SIMEX has the same properties as the regression calibration. It is easy to program SIMEX like regression calibration, and for small measurement error bias SIMEX might improve the estimates in some highly nonlinear models.

According to Carroll et al. (2006) an integral component of SIMEX is a self-contained simulation study resulting in graphical displays that illustrate the effect of measurement error on parameter estimates and the need for the bias correction. The graphical displays are useful when it is necessary to motivate or explain a measurement error model analysis. The bias due to the measurement error in almost any estimator of almost any parameter is readily estimated and corrected, at least approximately by SIMEX.

3.2.1 The Simulation Extrapolation Estimator for Linear Regression

SIMEX is a general-purpose technique for the measurement error correction. There are substantial literatures on the application and theory of SIMEX for purely parametric and nonparametric regression problems. It is a functional method, in the sense that it makes no assumptions about the distribution of the unobserved true covariate. A major strength of SIMEX is that it is extremely easy to implement: it requires only a program for computing estimates in the absence of measurement error, and is able to simulate additional measurement error being added to the process. The idea behind the method is most clearly seen in simple linear regression when the independent variable is subject to measurement error. Suppose, in the model

\[ Y = \beta_0 + \beta_1 X + \varepsilon, \]

the predictor \( X \) cannot be observed, but we observe \( W = X + U \) where \( U \) has mean zero and known variance \( \sigma_U^2 \). It is well known that the ordinary least squares estimate of the slope from regressing \( Y \) on \( W \) converges to

\[ \frac{\beta_1 \sigma_X^2}{\sigma_Y^2 + \sigma_U^2}, \]

where \( \sigma_X^2 \) denotes the variance of \( X \). The SIMEX method uses the relationship between the size of the measurement error which
is described by the measurement error variance $\sigma_U^2$ and the bias of the effect estimator when ignoring the measurement error. For any fixed $\lambda > 0$, suppose one repeatedly 'adds on' additional error with mean zero and variance $\lambda \sigma_U^2$ to $W$ via simulation to compute the ordinary least squares slope each time, and then takes the average. This simulation estimator consistently estimates

$$g(\lambda) = \frac{\beta_1 \sigma_x^2}{\sigma_x^2 + (1 + \lambda)\sigma_U^2}.$$

Since, formally at least, $g(-1) = \beta_1$, the idea is to plot $g(\lambda)$ against $\lambda \geq 0$, fit a model to this plot, and then extrapolate back to $\lambda = -1$. Here is the precise definition of the SIMEX estimator for the nonparametric regression. Suppose the original data is used to estimate $\beta_1$, which is $\hat{\beta}_{1,\text{naive}}$. There are $M - 1$ data sets available, each with successively larger measurement error variances, say, $\sigma_U^2 (1 + \lambda_m)$, where $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_M$ are known. The least square estimate of the slope of the $m^{th}$ data set is, $\hat{\beta}_{1,m}$, consistently estimates $\frac{\beta_1 \sigma_x^2}{\sigma_x^2 + (1 + \lambda_m)\sigma_U^2}$. Consider the data $(\lambda_m, \hat{\beta}_{1,m})$, dependent variable $\hat{\beta}_{1,m}$, and independent variable $\lambda_m$.

Asymptotically, the mean function of the regression has the form,

$$E(\hat{\beta}_{1,m} | \lambda) = g(\lambda) = \frac{\beta_1 \sigma_x^2}{\sigma_x^2 + (1 + \lambda)\sigma_U^2},$$

here $\lambda \geq 0$.

Now $g(-1) = \beta_1$, that is, the parameter of interest is obtained from $g(\lambda)$ by extrapolation to $\lambda = -1$. Actually, one would like error-free data sets and in terms of $\lambda_m$ this corresponds to having $\sigma_U^2 (1 + \lambda_m) = 0$ and thus $\lambda_m = -1$. 

3.2.2 The SIMEX Algorithm

To explain the SIMEX algorithm, Carroll et al. (2006) considered four combinations of measurement error models and measured data.

- In the first, a single measurement for each case is available and the measurement errors are homoscedastic with a known, or independently estimated variance.
- In the second, a single measurement for each case is available and the measurement errors are heteroscedastic with a known variance.
- In the third case, replicate measurements are assumed but no assumptions are made about the error variance. That is, it is not assumed that they are known and they could be homoscedastic or heteroscedastic.
- In the fourth case, it is shown that how the method is generalized to a certain multiplicative error model.

Case 1: Homoscedastic Errors with a Known Error Variance

SIMEX is a general methodology and is easy to understand when there is only one predictor $X$ subject to error and multiple covariates $Z$ measured without error. Suppose one observes $W_i = X_i + U_i$, where $U_i$ is a normal random variable with the variance $\sigma_U^2$ and is independent of $X_i$, $Z_i$ and $Y_i$. A minor violation of the assumption of normality of measurement errors is not critical in practice. It is assumed that the measurement error variance $\sigma_U^2$ is known or sufficiently well estimated to be regarded as known. Let $\Theta$ denotes the parameter of interest.

The first part of the algorithm is the simulation part; which involves simulations to create additional data sets of increasingly large measurement error $\sigma_U^2 \left(1 + \lambda \right)$. For any $\lambda \geq 0$, define $W_{m,i}(\lambda) = W_i + U_{m,i} \lambda^{\frac{1}{2}}$, $i = 1, 2, \ldots, n; m = 1, 2, \ldots, M$.
Hence $W_{m,j}(\lambda)$ is a remeasurement of $W_j$, and $W_j$ is a measurement of $X_i$. Note that,

$$Var \left( W_i \mid X_i \right) = \sigma^2_U$$

and

$$Var \left( W_{m,j}(\lambda) \mid X_i \right) = \sigma^2_U \left( 1 + \lambda \right) = \left( 1 + \lambda \right) Var \left( W_i \mid X_i \right)$$

(3.2.1)

The computer generated pseudo errors $\{U_{m,i}\}_{i=1}^n$ are mutually independent, identically distributed normal random variables with the mean 0 and variance $\sigma^2_U$. It is also independent of all the observed data.

The error variance in the remeasured data has been inflated by a multiplicative factor $(1 + \lambda)$ and in this case that equals zero when $\lambda = -1$.

Because $E \left( W_{m,j}(\lambda) \mid X_i \right) = X_i$, (3.2.1) implies that the mean squared error of $W_{m,j}(\lambda)$ as a measurement of $X_i$ is defined as,

$$MSE \left( W_{m,j}(\lambda) \right) = E \left[ \left( W_{m,j}(\lambda) - X_i \right)^2 \mid X_i \right] \rightarrow 0 \text{ as } \lambda \rightarrow -1.$$

This is the key property of the simulated pseudo or remeasured data. After generating the remeasured predictor, the corresponding naïve estimator is computed. Define $\hat{\Theta}_m(\lambda)$ be the estimator when $\{W_{m,j}(\lambda)\}_{j=1}^n$ are used, and compute the average of these estimators as,

$$\hat{\Theta}(\lambda) = M^{-1} \sum_{m=1}^M \hat{\Theta}_m(\lambda)$$

By design, $\hat{\Theta}(\lambda)$ is the sample mean of $\{\hat{\Theta}_m(\lambda)\}_{m=1}^M$ and hence the average of the estimates obtained from a large number of experiments with the same amount of measurement error. The reason for averaging over many simulations is that one might be interested in estimating extra bias due to added measurement error, and averaging reduces the Monte Carlo simulation variation. This is the simulation component of SIMEX.
The components of $\hat{\Theta}(\lambda)$ are all functions of the same scalar $\lambda$, and there is a separate extrapolation step for each component of $\hat{\Theta}(\lambda)$. The extrapolation step entails modeling each of the components $\hat{\Theta}(\lambda)$ as a function of $\lambda$ for $\lambda \geq 0$ and extrapolating the fitted model back to $\lambda = -1$. The vector of extrapolated values yields the simulation extrapolation estimator defined by $\hat{\Theta}$. The significance of $\lambda = -1$ follows from the fact that $\hat{\Theta}(\lambda)$ is calculated from the measurements having the variance,

$$Var\left(W_{m,i}(\lambda) \mid X_i\right) = \sigma^2_U \left(1 + \lambda\right)$$

and one wants to extrapolate the case in which the error variance in the measurements is zero, that is $\sigma^2_U \left(1 + \lambda\right) = 0$ or, equivalently, $\lambda = -1$.

Case 2: Heteroscedastic Errors with Known Error Variance

Suppose instead of observing $X_i$, one observes $W_i = X_i + U_i$, where $U_i$ is a normal random variable with a known variance $\sigma^2_U$, which is independent of $X_i$, $Z_i$, and $Y_i$. It is appropriate when $W_i$ is the mean of $k_i \geq 1$ replicate measurements, each has known variance $\sigma^2_U$, and $\sigma^2_{U,i} = \frac{\sigma^2_U}{k_i}$. In this algorithm the remeasurement procedure becomes

$$W_{m,i}(\lambda) = W_i + U_{m,i} \lambda^{\frac{1}{2}}, \quad i = 1, 2, \ldots, n, \quad m = 1, 2, \ldots, M.$$ 

The pseudo errors $\left\{U_{m,i}\right\}_{i=1}^n$ are mutually independent, independent of all the observed data and identically distributed, normal random variables with mean zero and variance $\sigma^2_{U,i}$. Here,
\[ \text{Var} \left( W_{m,i}(\lambda) \mid X_i \right) = \sigma_{U,i}^2 (1+\lambda) = (1+\lambda) \text{Var} \left( W_i \mid X_i \right), \text{and,} \ E \left( W_{m,i}(\lambda) \mid X_i \right) = X_i. \] So the two variances, \( \text{Var} \left( W_{m,i}(\lambda) \mid X_i \right) \), and \( \text{Var} \left( W_i \mid X_i \right) \), differ by a multiplicative factor, which vanishes when \( \lambda = -1 \). Consequently the mean squared error is

\[ \text{MSE} \left( W_{m,i}(\lambda) \right) = E \left[ \left( W_{m,i}(\lambda) - X_i \right)^2 \mid X_i \right] \to 0 \quad \text{as} \quad \lambda \to -1. \]

The average naïve estimate, \( \hat{\Theta}(\lambda) \), are calculated following the same way as of the homoscedastic error model. The SIMEX estimator \( \hat{\Theta}_{\text{SIMEX}} \) is defined by the modeling and extrapolating to \( \lambda = -1 \) as this is the value for which \( \sigma_{U,i}^2 (1+\lambda) = 0 \) for all \( i \).

Case 3: Heretoscedastic Error with Unknown Variances and Replicate Measurements

Consider an error model with arbitrary unknown heteroscedastic error variances. SIMEX estimation for this model was developed and studied by Devanarayan (1996) and Devanarayan and Stefanski (2002). For this model, \( k_i \geq 2 \) replicate measurements are necessary for each subjects in order to identify the error variances, \( \sigma_{U,i}^2 \). The assumed error model is

\[ W_{i,j} = X_i + U_{i,j}, \]

where \( U_{i,j} \) follows a Normal distribution with mean zero and variance \( \sigma_{U,i}^2 \), and independent of \( X_i, Z_i \) and \( Y_i \) with all \( \sigma_{U,i}^2 \) unknown.

With replicate measurements, the best measurement of \( X_i \) is the mean \( \bar{W}_{i,*} \), and for the measurement error analysis \( \{ Y_i, Z_i, W_i \}_{i=1}^m \) data is considered. The variance \( \sigma_{U,i}^2 \) is unknown, and the variance of the best measurement of \( X_i \) is inflated by the factor \( (1+\lambda) \); with replicate measurements, one can obtain variance inflated measurements by taking sub optimal linear combinations of the replicate measurements. This is done by the linear contrasts.
Let $C_{m,j} = (C_{m,j,1}, \ldots, C_{m,j,k_j})$ is a normalized contrast vector, where $\sum_j C_{m,j,j} = 0$, and

$$\sum_j C_{m,j,j}^2 = 1.$$ Define,

$$W_{m,j}(\lambda) = W_{m,j} + \frac{\lambda}{k_i} \sum_{j=1}^{k_i} C_{m,j}W_{i,j}$$

for $i = 1, 2, \ldots, n$, $m = 1, 2, \ldots, M$ and

$$E(W_{m,j}(\lambda) | X_i) = X_i$$

$$Var(W_{m,j}(\lambda) | X_i) = \frac{1}{k_i} \sigma^2_{U,j} (1 + \lambda) = (1 + \lambda) Var(W_i | X_i)$$

and the variances of the error in the remeasurements are inflated by a multiplicative factor that vanishes when $\lambda = -1$ and the mean squared error,

$$MSE(W_{m,j}(\lambda)) \rightarrow 0 \text{ as } \lambda \rightarrow -1.$$

Now by making contrasts random, we generate random, replicate version of (3.2.2) and find the average of $M$ remeasured data sets. The average naïve estimates $\hat{\Theta}(\lambda)$ are calculated exactly following the same way in previous two cases. Again, SIMEX estimator is obtained by modeling the relationship between $\hat{\Theta}(\lambda)$ and $\lambda$ and extrapolating to $\lambda = -1$

Case 4: Nonadditive Measurement Error

In the case of multiplicative error model, one can transfer it into additive by logarithmic transformation. Now the observed data error model is,

$$\log(W_i) = \log(X_i) + U_i,$$

Here $U_i$s are Normal (0, $\sigma^2_U$). The remeasured data are obtained as,

$$\log(W_{m,j}(\lambda)) = \log(W_i) + U_{m,j} \frac{1}{\lambda^2}.$$
Here $U_{m,i}$ are Normal $(0, \sigma_U^2)$ pseudo random variables. Consider the transformation,

$$ W_{m,i}(\lambda) = \exp \left\{ \log(W_i) + \sqrt{\lambda} U_{m,i} \right\} $$

Here the variance is increased by a multiplicative factor $(1 + \lambda)$ and this multiplier vanishes when $\lambda = -1$.

### 3.2.3 Example of SIMEX

Airborne contaminants occur in the gaseous form such as, gases and vapors. In scientific terminology, it is defined as a system of particles suspended in a gaseous medium, usually air in the context of occupational hygiene. Gases may exist in the form of airborne dusts, sprays, mists, smokes and fumes. In the occupational setting, all these forms may be important because they relate to a wide range of occupational diseases. Airborne dusts are of particular concern because they are well known to be associated with classical widespread occupational lung diseases such as the pneumoconioses, as well as with systemic intoxications such as lead poisoning, especially at higher levels of exposure. But, in the modern era, there is also an increasing interest in other dust-related diseases, such as cancer, asthma, bronchitis, allergic alveolitis, and irritation, as well as a whole range of non-respiratory illnesses, which may occur at much lower exposure levels.

We use a data set about chronic bronchitis and dust concentration of the German research foundation to illustrate the application of the SIMEX method. In this example, the comparison between the naïve approach and SIMEX method is also discussed. The dataset is part of an epidemiological study conducted by the German Research Foundation. The aim of the study is to examine the relationship between dust concentration at the working place and chronic bronchitis and to assess the threshold limiting value for the dust concentration. The data on 1246 workers were collected between 1960-1977 in a Munich plant.
The data set consists of the following four variables:

- **CBR**: Chronic Bronchitis Reaction 1=yes 0=no
- **Dust**: Dust concentration at the working place in mg/m^3
- **Smoking**: Does worker smoke? 1=yes 0=no
- **Years**: Duration of exposure in years

It is possible that the variable dust is subject to the measurement error. Because it is a continuous variable, the SIMEX-method is used here. Usually the measurement error variance is estimated by replication or by a validation study, which is not available in this case. For illustration we assume that the additive measurement error has a standard deviation $\sigma = 2$. The estimation by the SIMEX function are listed as follows.

**Table 3.3 Comparison between the Naïve and SIMEX Estimators**

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept</th>
<th>Dust</th>
<th>Smoking</th>
<th>Expo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>-3.04787</td>
<td>0.09189</td>
<td>0.67684</td>
<td>0.04015</td>
</tr>
<tr>
<td>SIMEX</td>
<td>-3.13978</td>
<td>0.12700</td>
<td>0.67829</td>
<td>0.03966</td>
</tr>
</tbody>
</table>

It can be seen from the displayed naïve estimates, there is nearly no measurement error correction for the estimates of the intercept, smoking and expo, while the corrected estimate for the variable dust differs substantially form the naïve estimator. Since the theory does not provide an optimal choice for the extrapolation function, different choices should be calculated and inspected graphically. For the analysis SIMEX package in R is used. The package **simex** features easy to use functions for correcting estimation in regression models with measurement error or misclassification via the SIMEX–method. It provides fast and easy ways to produce plots that illustrate the effect of measurement error or misclassification on parameters. Several additional
functions are available to help with various problems concerning misclassification or measurement error.

3.3 Idea of Score Function Method

The regression calibration and SIMEX are extensively applicable common methods for eliminating or reducing measurement error. These methods provide consistent estimators in linear regression, loglinear mean models, etc. The score function method is also widely applicable and provides fully consistent estimators. Consistency is attained by means of the fact that the estimators are M-estimators, whereas score functions are unbiased in the presence of measurement error. Consider the classical measurement error model, $W = X + U$, where $U$ is a normally distributed random vector with the mean zero and covariance matrix, $\Sigma_{UU}$. There are two broad classes of the score function methods;

- The conditional score method (Stefanski, Carroll, 1987) exploits special structures in important models such as linear, logistic, poisson, log linear and gamma-inverse regression, by using a traditional statistical device, and conditioning on sufficient statistics, to obtain estimators.
- The corrected score method effectively estimates the estimator that one would use if there are no measurement error, see Carroll et al. (2006).

3.3.1 Linear Regression Corrected and Conditional Scores

Let us consider the multiple linear regression models with the mean and variance, $E(Y \mid X, Z) = \beta_0 + \beta_1' X + \beta_2' Z$ and $\sigma^2$ respectively. Under the classical additive measurement error model, one observes $W = X + U$ instead of $X$, where $U$ is a normally distributed random vector with the mean zero and covariance matrix, $\Sigma_{UU}$. Rewrite the regression parameters as $\Theta_1 = (\beta_0, \beta_1', \beta_2')'$ and $\Theta_2 = \sigma^2$ and $\Theta = (\Theta_1, \Theta_2)'$. The ordinary least square score function for multiple linear regression when there is no measurement error is
The normal equations, \( \{ Y_i - (1, X'_i, Z'_i) \Theta_1 \} \begin{pmatrix} 1 \\ X_i \\ Z_i \end{pmatrix} \), \( i = 1, 2, \ldots, n \), are the least squares score functions for the regression parameters \( \Theta_1 \), and the dimension of \( \Theta_1 \) is \( p \).

### 3.3.1.1 Linear Regression Conditional Score

This approach requires no assumption about the error prone predictor \( X \), but the derivation of the method assumes normality of the true regression equation error \( U_i \). Let,

\[
\Delta_i = W_i + \frac{Y_i}{\sigma^2} \Sigma_{U_i} \beta_i
\]  

(3.3.2)

Given \( X_i, Z_i \), the random variables \( Y_i \) and \( \Delta_i \) are linear functions of jointly normal random vectors and thus are jointly normal, conditionally on \((X_i, Z_i)\). Therefore, the conditional distribution of \( Y_i \) given \((X_i, Z_i, \Delta_i)\) is also normal and the standard multivariate normal calculation shows that

\[
E \left( Y_i \mid X_i, Z_i, \Delta_i \right) = E \left( Y_i \mid Z_i, \Delta_i \right) = \frac{\beta_{0i} + \beta_{1i} \Delta_i + \beta_{2i} Z_i}{1 + \frac{1}{\sigma^2} \beta_{0i} \Sigma_{U_i} \beta_i} 
\]  

(3.3.3)

The conditional score becomes,
\[
\Lambda_{\text{Cond}}(Y_i, Z_i, W_i, \Theta) = \left[ \begin{array}{c}
\{ Y_i - E(Y_i | Z_i, \Delta_i) \} \begin{bmatrix} 1 \\ Z_i \\ \Delta_i \end{bmatrix} \\
\sigma^2 \left( \frac{Y_i - E(Y_i | Z_i, \Delta_i))^2}{\text{Var}(Y_i | Z_i, \Delta_i)} \right) \end{array} \right]
\]

Hence, \( E[\Lambda_{\text{Cond}}(Y_i, Z_i, W_i, \Theta) | Z_i, \Delta_i] = 0 \). Therefore, \( \Lambda_{\text{Cond}} \) can be used to form unbiased estimating equations, \( \sum_{i=1}^{n} \Lambda_{\text{Cond}}(Y_i, Z_i, W_i, \Theta) = 0 \). But in practice, one can estimate the parameters by solving the small sample modified estimating equations

\[
\sum_{i=1}^{n} \left[ \begin{array}{c}
\{ Y_i - E(Y_i | Z_i, \Delta_i) \} \begin{bmatrix} 1 \\ Z_i \\ \Delta_i \end{bmatrix} \\
\frac{n - p}{n} \sigma^2 \left( \frac{Y_i - E(Y_i | Z_i, \Delta_i))^2}{\text{Var}(Y_i | Z_i, \Delta_i)} \right) \end{array} \right] = 0
\]

### 3.3.1.2 Linear Regression Corrected Score

The general method of constructing corrected scores uses complex variable and complex valued functions. In this case, computer generates the random variables to help in defining an estimator and the random variables are defined as follows.

For \( m = 1, 2, \ldots, M \), generate random variables \( U_{m,i} \) that are independent normal vectors with the mean zero and covariance matrix, \( \Sigma_{UU} \). Consider the complex valued random variable,

\[
\tilde{W}_{m,i} = W_i + l U_{m,i} \quad \text{where, } l = \sqrt{-1}.
\]
The Monte Carlo corrected score is obtained as follows,

i. Replace $X_i$ by $\tilde{W}_{m,i}$ in (3.3.1).

ii. Take the real part of the expression to eliminate the imaginary part.

iii. Average over multiple sets of pseudo random vectors, where, $m = 1, 2, \ldots, M$.

For linear regression,

$$\Lambda_{MCCS,M}(Y_i, Z_i, W_i, \Theta) = \frac{1}{M} \sum_{m=1}^{M} \text{Re} \Lambda_{LS}(Y_i, Z_i, \tilde{W}_{m,i}, \Theta)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{n-P}{n} \sigma^2 - \{Y_i - (1, Z'_i, W'_i)\theta_1 \}^2 + \beta'_i \hat{M}_{U,i} \beta_1 \right]$$

where, $\hat{M}_{U,i} = \frac{1}{M} \sum_{m=1}^{M} U_{m,j} U'_{m,i}$, and $E(\hat{M}_{U,i}) = \Sigma_U$.

The corrected score $\Lambda_{MCCS,M}(Y_i, Z_i, W_i, \Theta)$ is an unbiased estimator of the score that would have been be used for $\Lambda_{LS}(Y_i, X_i, Z_i, \Theta)$, if the measurement error were not present. Again, under the regularity condition, the estimating equations,

$$\sum_{i=1}^{n} \Lambda_{MCCS,M}(Y_i, Z_i, W_i, \Theta) = 0 \quad (3.3.4)$$

is a consistent and asymptotically normal sequence of solutions. Simplifying (3.3.4), the corrected scores for linear regression becomes,

$$\hat{\Theta}_1 = (\hat{M}_{1ZW,1ZW} - \hat{\Omega})^{-1} \hat{M}_{1YW}$$
\[ \hat{\sigma}^2 = (n - p)^{-1} \sum_{i=1}^{n} \{ (Y_i - \hat{Y}_i)^2 - \hat{\beta}_i \hat{\Sigma}_{UU} \hat{\beta}_i \} \]

where, 
\[ \hat{M}_{1ZW,1ZW} = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 & Z_i' & W_i' \\ Z_i & Z_i' & Z_i W_i' \\ W_i & W_i' & W_i W_i' \end{pmatrix}, \quad \hat{\Theta}_i = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Sigma_{UU} \end{pmatrix} \]
\[ \hat{\Sigma}_{UU} = \frac{1}{n} \sum_{i=1}^{n} \hat{M}_{U,i}, \]
\[ \hat{M}_{Y,1ZW} = \frac{1}{n} \sum_{i=1}^{n} Y_i \begin{pmatrix} 1 \\ Z_i' \\ W_i' \end{pmatrix} \]
and \( \hat{\Sigma}_{UU} = \left( \hat{\Sigma}_{UU} + \hat{\Sigma}_{U} \right) \hat{\Theta}_i. \)

Since \( \hat{\Sigma}_{UU} \) converges to \( \Sigma_{UU} \) as \( M \to \infty \), we substitute \( \hat{\Sigma}_{UU} \) for \( \Sigma_{UU} \) as an estimate of the measurement error bias. For many measurement error models, substituting the complex variate \( \tilde{W}_{m,i} \) for \( X_i \) into a score function which is unbiased in the absence of measurement error, taking real part and averaging over \( m = 1, 2, \ldots, M \) results in an unbiased score that is a function of the observed data. For the corrected score method, no assumptions are made about the true regression equation error, \( \epsilon_i = Y_i - E(Y_i | X_i, Z_i) \) and the unobserved predictor \( X \).

### 3.3.2 Logistic Regression Corrected and Conditional Scores

Consider multiple logistic regression models,
\[ \Pr(Y = 1 | X, Z) = G \left( \beta_0 + \beta_1 X + \beta_2 Z \right) \]

Here \( G(\nu) = \frac{1}{1 + \exp(-\nu)} \). Suppose instead of observing \( X \), one observes \( W = X + U \) with \( U \) following \( N(0, \Sigma_{UU}) \), where \( \Sigma_{UU} \) is known. Let, \( \Theta_i = (\beta_0, \beta_1, \beta_2)' \). The maximum likelihood score function for multiple regressions when there is no measurement error is,
\[ A_{ML}(Y_i, X_i, Z_i, \Theta) = \begin{pmatrix} Y_i - G \left( (1, X_i', Z_i')' \Theta \right) \\ X_i \\ Z_i \end{pmatrix} \]
3.3.2.1 Logistic Regression Conditional Scores

For the conditional score method of logistic regression, define \( \Delta_i = W_i + Y_i \Sigma_{UL} \beta_i \).

Conditioned on \((X_i, Z_i)\), both \(Y_i\) and \(W_i\) have exponential family densities. Exponential family calculations show that,

\[
E \left( Y_i \mid X_i, Z_i, \Delta_i \right) = G \left( \beta_0 + \beta_1 \Delta_i + \beta_2 Z_i - \frac{1}{2} \beta_i^T \Sigma_{UL} \beta_i \right)
\]

\[= E \left( Y_i \mid Z_i, \Delta_i \right) = \Pr \left( Y = 1 \mid Z_i, \Delta_i \right) \]

The conditional distribution of \(Y_i\) given \((Z_i, \Delta_i)\) does not depend on \(X_i\) and the conditional becomes \( \Lambda_{\text{Cond}} \left( Y_i, Z_i, W_i, \Theta \right) = \left[ \begin{array}{c} \{ Y_i - E \left( Y_i \mid Z_i, \Delta_i \right) \} \\ Z_i \end{array} \right] \) which has the property

\[
E \left[ \Lambda_{\text{Cond}} \left( Y_i, Z_i, W_i, \Theta \right) \mid Z_i, \Delta_i \right] = 0.
\]

That is, its conditional mean vanishes. Therefore, \( \Lambda_{\text{Cond}} \) can be used to form unbiased estimating equations \( \sum_{i = 1}^{n} \Lambda_{\text{Cond}} \left( Y_i, Z_i, W_i, \Theta \right) = 0. \)

3.3.2.2 Logistic Regression Corrected Scores

The logistic model does not satisfy the smoothness conditions required by the corrected score theory; but corrected score method can be applied when the measurement error variance is not large. This produces nearly consistent estimator. The corrected score for logistic regression is

\[
\Lambda_{\text{MCCS,M}} \left( Y_i, Z_i, W_i, \Theta \right) = \frac{1}{M} \sum_{m=1}^{M} \Re \Lambda_{\text{LS}} \left( Y_i, Z_i, \widetilde{W}_{m,i}, \Theta \right)
\]

Hence the solution of the resulting expression may not have a closed form.
3.4 Conditional Score Function: Basic Theory

Canonical generalized linear models for $Y$ given $(X, Z)$ have the following density or the mass function,

$$f(y | x, z, \Theta) = \exp \left\{ \frac{y\eta - A(\eta)}{\Phi} + c(y, \Phi) \right\},$$

where $\eta = \beta_0 + \beta'_1 X + \beta'_2 Z$ is called the natural parameters and $\Theta = (\beta_0, \beta'_1, \beta'_2, \Phi)$ be the unknown parameters to be estimated. The mean and the variance of $y$ are $A'(\eta)$ and $A''(\eta)$ respectively. This class of models includes;

i. Linear Regression: Mean=$\eta$, Variance=$\Phi$, $A(\eta) = \frac{\eta^2}{2}$,

$$c(y, \Phi) = -\frac{y^2}{2\Phi} - \log \sqrt{2\pi \Phi}.$$

ii. Logistic Regression: Mean=$G(\eta)$, Variance=$G'(\eta)$, $\Phi = 1$,

$$A(\eta) = -\log \{1 - G(\eta)\}, c(y, \Phi) = 0. \text{ Hence } G(\eta) = \frac{1}{1 + \exp(-\eta)}.$$

iii. Poisson Log linear Regression: Mean=$\exp(\eta)$, Variance=$\exp(\eta)$, $\Phi = 1$,

$$A(\eta) = \exp(\eta); c(y, \Phi) = -\log(y!).$$

iv. Gamma Inverse Regression: Mean=$-\frac{1}{\eta}$, Variance=$-\frac{\Phi}{\eta}$, $A(\eta) = -\log(-\eta)$,

$$c(y, \Phi) = \frac{1}{\Phi} \log \left( \frac{y}{\Phi} \right) - \log \left( \frac{y}{\Phi} \right)^{1/\Phi}.$$

Now, if $X_i$ were observed, then $\Theta$ is estimated by solving

$$\sum_{i=1}^{n} \Lambda_{QL} (Y_i, X_i, Z_i) = 0 \quad (3.4.1)$$
Subsequently, \( \eta_i = \beta_0 + \beta_1^i X + \beta_2^i Z , i = 1, 2, \ldots, n \). For certain models (3.4.1), produces maximum likelihood estimators.

### 3.5 Corrected Score Function: Basic Theory

The corrected score function is applied to any model for which the usual estimator in the absence of measurement error is an M-estimator. The basic idea is following;

i. Let \( \Lambda_{\text{true}} (Y, X, Z, \Theta) \) denote the score function that would be used for estimation if \( X \) were observed. This could be a nonlinear least square score or a likelihood score or so on.

ii. Since \( X \) is not observed, \( \Lambda_{\text{true}} (Y, X, Z, \Theta) \) cannot be used for estimation; therefore, one needs to construct an unbiased estimator of \( \Lambda_{\text{true}} (Y, X, Z, \Theta) \) based on observed data. Let the new score function be, \( \Lambda_{CS} (Y, Z, W, \Theta) \) which has the property,

\[
E \left( \Lambda_{CS} (Y, Z, W, \Theta) \right) = \Lambda_{\text{true}} (Y, X, Z, \Theta).
\]

iii. The corrected score function, \( \Lambda_{CS} (Y, Z, W, \Theta) \) is used for estimation for \( \Theta \), the calculation of the standard error, inference using M-estimator techniques.

Since there is no assumption about \( X \), the corrected score method provides an attractive approach to the consistent estimation when data are measured with error.
3.6 Comparison of Conditional and Corrected Score Functions

Conditional score and corrected score methods are both functional methods and neither one requires an assumption on $X$ for consistency to hold in general. But they differ in other important areas.

Conditional scores are derived under specific assumptions about both the model for $Y$ given $(X, Z)$ and the error model for $W$ given $X$ whereas corrected scores assume only a correct estimating function if $X$ were observed, and sufficient assumptions on the error model enable unbiased estimation of the true-data estimating function. If the assumptions of the conditional score method are satisfied, it will usually be more efficient. Sometimes the conditional score requires numerical summation and integration. The exact corrected method also needs integration; but Monte Carlo corrected score method come with a simple built in solution to the computational problem when the required integrals are not analytically tractable. Moreover, conditional score method and some of its extensions have a theoretical advantage in terms of efficiency.

In addition, both conditional score and corrected score have further extensions. Conditional score method can be easily extended to the case that the model for $W$ given $X$ is a canonical generalized linear model with natural parameter $X$. Buzas and Stefanski (1996) showed a simple extension of corrected score method to additive non-normal error models when the true data score function depends on $X$ only through $\exp(\beta' X)$ and the measurement error possesses a moment generating function.

3.7 Comparison of Different Approaches Using Real Life Data

This section presents the comparisons among the existing approaches for measurement error model. The real life data set is used for this purpose, and it is collected from Fuller (2006).
3.7.1 Application of the Method to Simple Linear Regression Model

To illustrate the application of the existing method for simple linear regression model wherein the single predictor variable is subject to error, we consider the small data set on yield of corn and available soil nitrogen adapted from Voss (1969) and presented by DeGracie and Fuller (1972). The comparisons among the existing methods are also shown.

Nitrogen (N) is typically one of the largest corn fertilization expenses. Nitrogen application is critical because it significantly improves corn yield in many crop rotations. When choosing N rates, producers need to carefully consider both achieving most profitable economic return and advancing environmental stewardship. The data given in Table 3.4 are yields of corn and determination of available soil nitrogen collected at 11 sites on Marshall soil in Iowa. The experimental measurements done by Voss (1969) included soil test values for nitrogen, phosphorous, and potassium in the surface soil and yield of the corn crop. The estimates of the soil nitrogen contain measurement error arising from two sources; first only a small sample of soil is selected from each plot and as a result, there is sampling error associated with the use of sample to represent the whole; second there is a measurement error associated with the chemical analysis used in the laboratory to determine the level of nitrogen in the soil sample. The variance arising from these two sources has been estimated to be 57 and for the next analysis we will assume that the measurement error variance is known.
Table 3.4 Yield of Corn on Marshall Soil in Iowa

<table>
<thead>
<tr>
<th>Site</th>
<th>Yield ($Y$)</th>
<th>Soil Nitrogen ($W$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>115</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>86</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>110</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>91</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>70</td>
</tr>
<tr>
<td>9</td>
<td>99</td>
<td>94</td>
</tr>
<tr>
<td>10</td>
<td>104</td>
<td>69</td>
</tr>
<tr>
<td>11</td>
<td>96</td>
<td>51</td>
</tr>
</tbody>
</table>
Figure 3.2 Scatter plot of soil nitrogen and yield of corn

It is becoming increasingly common for precision agriculture service providers to create scatter plots and calculate bi-variate regression correlation coefficients for paired data. When this is applied to soil nitrogen and yield data sets, as shown in Figure 3.2, the results typically show statistically significant correlation. The yield and soil nitrogen from this Iowa cornfield has a statistically significant (at the 1% significance level) correlation coefficient of 0.64153. The correlation analysis is an important first step in investigating the causes of yield variability. The visual picture of the scatter plot, the statistical correlation of the data indicates that the patterns are not random.

To analysis the data, consider the classical linear regression model with one independent variable, soil nitrogen, defined as
\[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]

Since we cannot observe the exact amount of soil nitrogen, \( x_i \) but observe an estimate of \( x_i \), we represent the observed nitrogen by \( w_i \), where \( w_i \) satisfies
\[ w_i = x_i + u_i \]

and \( u_i \) is the measurement error introduced by sampling and laboratory analysis. Under the assumption, \( (x_i, \epsilon_i, u_i) \sim \mathcal{N}(\mu_x, 0, 0), \text{diag}(\sigma_{xx}, \sigma_{\epsilon\epsilon}, \sigma_{uu}) \) from (1.1.8), (1.1.9) the naïve least square estimate of \( \beta_1 \) becomes,
\[ \hat{\beta}_1 = \frac{\hat{\beta}_{1,LS}}{\hat{\rho}} \]

Here,
\[ \hat{\rho} = \frac{s^2_x}{s^2_x + s^2_u} \]

Here \( \hat{\rho} \) is the estimated reliability factor, and \( s^2 \) denotes the sample variance of the subscript variable.

Table 3.5 Comparison of Estimators: Naïve, Regression Calibration, SIMEX for the Simple Linear Regression Model with Single Variable Measured with Error

<table>
<thead>
<tr>
<th>Method of estimating regression parameters</th>
<th>Estimate of ( \beta_0 ), ( \hat{\beta}_0 )</th>
<th>Estimate of ( \beta_1 ), ( \hat{\beta}_1 )</th>
<th>Standard error ( \left( \hat{\beta}_0 \right) )</th>
<th>Standard error ( \left( \hat{\beta}_1 \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Method</td>
<td>73.1529</td>
<td>0.3440</td>
<td>9.9513</td>
<td>0.1371</td>
</tr>
<tr>
<td>Regression Calibration</td>
<td>67.564</td>
<td>0.42316</td>
<td>12.5423</td>
<td>0.1745</td>
</tr>
<tr>
<td>SIMEX</td>
<td>70.5125</td>
<td>0.3811</td>
<td>10.2966</td>
<td>0.1406</td>
</tr>
</tbody>
</table>

The Table 3.5 presents the estimated slope and intercept of the yield of corn and soil nitrogen data obtained by different methods. The estimates of the slopes for this simple linear
regression model are almost similar (0.3440, 0.42316, 0.3811) in all the approaches; but the estimates of the intercepts differ substantially. The same thing appears for the standard errors of the slope and intercept.
CHAPTER 4

ESTIMATION OF ERROR-PRONE REGRESSION PARAMETERS USING QVF AND EMPIRICAL LIKELIHOOD METHOD

The statistical consideration of models containing measurement errors began as early as 1877; see Fuller (1987, p.30). Most of the past work has been done on univariate linear model with constant error variances; more recently, work has been done on multivariate, nonlinear, and non-constant error variance models. There is a large literature in all branches of science concerned with inference in linear and nonlinear models with measurement errors. A review of the commonly used techniques can be found in Fuller (1987) and Carroll, Ruppert, and Stefanski (1995). Regression work on the classical independent additive measurement error model dates back to Frisch (1934), who derived bounds on the slope and the constant term in linear regression with measurement error. The method of instrumental variables (IV) is a popular method for obtaining consistent estimators of the parameters of interest in linear models with classical independent additive measurement error. In nonlinear regression models, Hausman, Ichimura, Newey, and Powell (1991), and Hausman, Newey, and Powell (1995) generalized this IV method to polynomial functions in the presence of double measurements on the error prone variables. Li (2002) and Schennach (2004) presented methods for nonlinear regression models with classical measurement error and double measurements. Hong and Tamer (2003) and Taupin (2001) use distributional assumptions on the measurement error to obtain a simple estimator in nonlinear models when no auxiliary data is present. Chesher (1991) presented useful approximation methods to the true distribution and parameters of interest. Most of these research impose the classical errors in variables assumption. On the other hand, Horowitz and Manski (1995) used a different model of measurement error where they assume that the observed sample
is contaminated or corrupted. Other papers allowing for non-classical measurement errors are the ones assuming the presence of true validation data; see Carroll and Wand (1991), Sepanski and Carroll (1993) and Lee and Sepanski (1995). Carroll and Wand (1991) use a semi parametric maximum likelihood estimator in a logistic regression model with covariate measurement error. Sepanski and Carroll (1993) use a quasi likelihood approach to estimate nonlinear regression models. In addition to the existence of a validation sample, both works assumed that the conditional distribution of the response given the mismeasured variable and the true variable is the same as the conditional distribution given the true variable. Lee and Sepanski (1995) proposed an innovative estimator for nonlinear regression problems in the presence of validation data. Their method uses a least squares projection onto a fixed finite dimensional collection of functions as a replacement for the conditional expectation of the nonlinear function.

4.1 Estimation of Regression Parameters

The effects of measurement error have long been recognized. In many studies the predictor variables can only be conducted with substantial measurement error. For the standard regression model with explanatory variable measured with error, it is well known that the estimated regression coefficients, on average, are attenuated toward the origin. Bias of this nature is commonly referred to as attenuation or attenuation to the null. In fact, though, even this simple conclusion has to be qualified, because it depends on the relationship between the measurement, $W$ and the true predictor $X$, and possibly other variables in the regression model. In particular, the effect of measurement error depends upon the model under consideration and on the joint distribution of the measurement error and the other variables. In multiple linear regression, the effects of measurement error may vary depending on; see Gail and Bénichou (2000):

i. The regression model, be it additive or multiplicative model
ii. Whether or not predictor measured with error is univariate or multivariate and,

iii. The presence of bias in the measurement.

The effects may range from simple attenuation described above to situations where the real effects are hidden, observed data exhibit relationship that are not present in the error free data and even the sign of the estimated coefficients are reversed relative to the case with no measurement error. Many epidemiological studies include the investigations of the association between the changes in a risk factor, like blood pressure, cholesterol level and some adverse outcome, such as severity of the disease. Typically, these predictor variables, risk factors are measured with error. In the previous chapter we discuss the existing methods for measurement error bias correction; the regression calibration, SIMEX and the score function methods. In this chapter we are going to discuss a computational based approach; to estimate the unknown regression parameters of the error prone variables, we propose the quasi-likelihood variance function method which was introduced by Wedderburn (1974), then we apply empirical likelihood method to obtain the estimate of the regression coefficients. The quasi-likelihood variance function is very useful in the sense that it requires assumptions only on the first two moments, rather than the entire distribution of the data. The quasi-likelihood variance function approach is practical because in many situations the exact distribution of the observations is unknown. Moreover, this quasi-likelihood variance function has statistical properties similar to those of a log-likelihood function.

4.1.1 Simple Linear Regression Model with Measurement Error

Let \( y_i, x_i, \) and \( w_i \) represent the response, the true explanatory variable and its measurement, respectively, for the \( i \)th observation in a sample of size \( n \). For now, suppose that \( x_i \) is one dimensional and there is no additional explanatory variable. Assume,
\[ y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (4.1.1) \]

\[ w_i = x_i + u_i \]

At this point, \( i = 1, \ldots, n \).

Furthermore assume, \( (x_i, \varepsilon_i, u_i) \overset{\text{i.i.d.}}{\sim} N \left( \mu_x, 0,0 \right), \text{diag} \left( \sigma_{xx}, \sigma_{z}, \sigma_{uu} \right) \). The observed measure of \( x_i \) is \( w_i \), may be obtained by asking people question, by reading imperfect instrument, or by performing a laboratory analysis. It is assumed that the variance of the measurement error \( \sigma_{uu} \) has been determined by making a large number of independent replicated measurements. The estimators of the remaining parameters are derived under the assumption that \( \sigma_{uu} \) is known.

Since \( V_i = (Y_i, W_i) \) is bivariate normal, the sample mean, \( \bar{V} = (\bar{Y}, \bar{W}) \) and sample covariances are, \( (m_{YY}, m_{YW}, m_{WW}) \), where, \[ m_{WW} = \frac{1}{n-1} \sum_{i=1}^{n} (W_i - \bar{W})(Y_i - \bar{Y}) \] form a set of sufficient statistics for the estimation of the parameters. If the parameter vector is identified, the maximum likelihood estimator will be a function of these statistics. If there are no parametric restrictions on the covariance matrix of \( V_i \), then \[ \frac{1}{n} \sum_{i=1}^{n} (V_i - \bar{V}) (V_i - \bar{V}) \] is the maximum likelihood estimator of the covariance matrix of \( V_i \). Now under the model, (4.1.1), the population moments of \( V_i = (Y_i, W_i) \) satisfy,

\[ (\sigma_{YY}, \sigma_{YW}, \sigma_{WW}) = (\beta_1^2 \sigma_{xx} + \sigma_{z}, \beta_1 \sigma_{xx}, \sigma_{xx} + \sigma_{zu}) \]

and

\[ (\mu_Y, \mu_W) = (\beta_0 + \beta_1 \mu_x, \mu_x) \quad (4.1.2) \]
We create estimators of the unknown parameters by replacing the unknown population moments on the left side of (4.1.2) with their sample estimators to obtain a system of equations in the unknown parameters. Solving, we have,

\[
\hat{\beta}_i = (m_{wW} - \sigma_{uu})^{-1}m_{wY} \\
(\hat{\sigma}_{xx}, \hat{\sigma}_{zx}) = (m_{wW} - \sigma_{uu}, m_{YY} - \hat{\beta}_i m_{wY}) \\
\left(\hat{\mu}_x, \hat{\beta}_0\right) = (\hat{W}, \bar{Y} - \hat{\beta}_i \hat{W})
\]

(4.1.3)

The estimators of \(\hat{\sigma}_{xx}\) and \(\hat{\sigma}_{uu}\) in (4.1.3) will be positive if and only if,

\[
m_{YY} (m_{wW} - \sigma_{uu}) - m_{wY}^2 > 0
\]

(4.1.4)

Estimators of samples in which (4.1.3) is violated are \(\hat{\sigma}_{zx} = 0\), \(\hat{\beta}_i = m_{wY}^{-1} m_{YY}\) and \(\hat{\sigma}_{xx} = m_{yy}^{-1} m_{wY}^2\) with \(\left(\hat{\mu}_x, \hat{\beta}_0\right) = (\hat{W}, \bar{Y} - \hat{\beta}_i \hat{W})\).

4.1.2 Estimation of Simple Linear Regression Parameter Using QVF and Empirical Likelihood (EL) Method

This section introduces a measurement error bias correction approach in simple linear regression model based on quasi likelihood and variance function (QVF) and empirical likelihood method. Later a general application of this approach will be discussed.

In the regression model, the data are of the form \((y_i, x_i)\) for \(1 \leq i \leq n\). Here \(x_i\) is the observed value for the \(i\)th observation in a sample of size \(n\) and \(y_i \in IR\); that is, \(x_i\) contains explanatory variable on the \(i\)th case for which \(y_i\) is the response. Instead of observing \(x_i\), one observes \(w_i\). The response \(y_i\) is the observed value of,

\[
y_i = \beta_0 + \beta_1 x_i + \varepsilon_i
\]

(4.1.5)
At this time, $\varepsilon_i$ is a random variable with mean 0 and variance $\sigma^2(x_i) < \infty$. In addition, the $n$ random variables $\varepsilon_i$ are independent. The quasi likelihood and variance function technique we have to model the mean and the variance function of the response, not its entire distribution.

Suppose, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$. Therefore, for simple linear regression model we postulate,

$$E(Y \mid X) = \mu(X, \beta)$$

and,

$$\text{Var}(Y \mid X) = \sigma^2 \nu^2(X, \beta, \theta) \quad (4.1.6)$$

This requires the computation of the mean and the variance function of the observed data, which is given by,

$$E(Y \mid W) = E\{\mu(X, \beta) \mid W\} \quad (4.1.7)$$

$$\text{Var}(Y \mid W) = \sigma^2 E\{\nu^2(X, \beta, \theta) \mid W\} + \text{Var}\{\mu(X, \beta) \mid W\} \quad (4.1.8)$$

Assume that replication or instrumental data are available to give an estimate $\hat{x}_i$ for $E(x_i \mid w_i)$. Replacing $x_i$ by $\hat{x}_i$ in (4.1.6) establishes an approximating QVF measurement error model as follows,

$$E(y_i \mid w_i) = \mu(\hat{x}_i, \beta) \quad (4.1.9)$$

$$\text{Var}(y_i \mid w_i) = \sigma^2 \nu^2(\hat{x}_i, \beta, \theta) \quad (4.1.10)$$

Specification of the mean and variance of the models, (4.1.9) and (4.1.10) allow constructing the estimating equations based on quasi likelihood function:

$$\sum_{i=1}^{n} \phi(y_i, w_i, \beta \mid \sigma, \theta) \mu'(\hat{x}_i, \beta) = 0 \quad (4.1.11)$$
Accordingly, $\mu' \hat{x}_i, \beta$ is the column vector of the derivatives $\mu(\cdot)$ with respect to $\beta$; and

$$
\phi(y_i, w_i, \beta | \sigma, \theta) = \frac{y_i - \mu(\hat{x}_i, \beta)}{\sigma^2 v^2(\hat{x}_i, \beta, \theta)}
$$

Suppose the estimate of the nuisance parameters are obtained using least square estimate; (Seber, 1977), $\hat{\sigma}, \hat{\theta}$. At this time, using this estimate construct an auxiliary variable,

$$
M_i(\beta^*) = \phi(y_i, w_i, \beta | \hat{\sigma}, \hat{\theta}) \mu'_{\text{sub}}(\hat{x}_i, \beta)
$$

Similarly, $\mu'_{\text{sub}}(\hat{x}_i, \beta)$ is the column vector of the derivatives of $\mu(\cdot)$ with respect to $\beta$. By construction $M_i(\beta^*)$ are independently and identically distributed.

Let

$$
p_i = \Pr(v = y_i, w = w_i), \quad i = 1, \ldots, n
$$

Followed by (4.1.13) the log empirical likelihood function becomes,

$$
l_n = \sum_{i=1}^{n} \log(p_i)
$$

The optimal computational problem is to maximize (4.1.14) over $p_i$ subject to the following constrains,

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

$$
\sum_{i=1}^{n} p_i M_i(\beta^*) = 0
$$

At this instant, using Lagrange multiplier technique, the optimal $p_i$’s are given by,

$$
\hat{p}_i = \frac{1}{n} \frac{1}{1 + \lambda' M_i(\beta^*)}, \quad i = 1, \ldots, n
$$

The Lagrange multiplier $\lambda$ solves the following equation,
\[
\sum_{i=1}^{n} \frac{M_i(\beta^*)}{1 + \lambda^* M_i(\beta^*)} = 0
\] (4.1.17)

Subsequently, the value of \( \lambda \) may be found by numerical search.

Therefore, the likelihood ratio statistics for the inference of \( \beta \) is,

\[
R_n(\beta) = -2 \sum_{i=1}^{n} \log \left(n \hat{p}_i \right)
\] (4.1.18)

By following Wilks’ theorem, \( R_n(\beta) \) has a limiting distribution \( \chi^2 \) for this simple linear regression model and a 100 \((1 - \alpha)\)% empirical likelihood confidence interval is formed by taking those values \( \beta \) for which \( R_n(\beta) \leq \chi^2(1-\alpha) \). For this reason, an asymptotic 100 \((1 - \alpha)\)% confidence region for \( \beta \) is given by,

\[
S(\alpha) = \{ \beta : R_n(\beta) \geq \chi^2(\alpha) \}
\]

At this point, \( \chi^2(\alpha) \) is upper 100 \( \alpha \)% quantile of the \( \chi^2 \) distribution.

4.2 Multiple Regression: Single Covariate Measured with Error

In multiple linear regression, the effects of measurement error are more complicated, even for the classical additive error. Consider the case, where \( X \) is a scalar, but there are additional predictors \( Z \) measured without error. In the linear model, the mean is \( \beta_0 + \beta_X X + \beta_Z Z \) under the usual conditions of independence of errors, the least squared regression estimator of the coefficient of \( W \), which is observed instead of \( X \), consistently estimates \( \rho \beta_X \); where,

\[
\rho = \frac{\sigma_{X|ZW}^2}{\sigma_{W|Z}^2} = \frac{\sigma_{X|Z}^2}{\sigma_{X|Z}^2 + \sigma_U^2}
\] (4.2.1)
Here, $\sigma^2_{w|x}$ and $\sigma^2_{x|z}$ are the variances of the regressions of $W$ on $Z$ and $X$ on $Z$ respectively.

The reliability factor $\rho$ is equal to the simple linear regression attenuation (1.1.10),

$$\rho = \frac{\sigma^2_{x}}{\sigma^2_{x} + \sigma^2_{u}},$$

only when $X$ and $Z$ are uncorrelated. The basic point is that the attenuation depends on the relationship among the predictors. This result has important consequences in many studies, like in epidemiology when the interest centers on the effects of covariates measured without error. Consider the case when $Z$ is a binary exposure variable, which is classified correctly and $X$ is an important predictor measured with significant error, then Carroll (2006) shows that ignoring measurement error produces a consistent estimate of the true exposure effect only if the design is balanced; that is $X$ has the same mean in both groups and is independent of treatment. With considerable imbalance, the naïve analysis may lead to the conclusion that:

a. There is a treatment effect when none actually exists and

b. The effects are negative when they are actually positive or vice versa.

That is naïve analysis may produce very misleading results.

4.2.1 Estimation of Multiple Linear Regression Parameter Using QVF and Empirical Likelihood (EL) Method

In this section, a generalization of the simple linear regression model with single predictor measured with error is considered and the measurement error bias correction method based on QVF and empirical likelihood method is discussed. Adopt the multiple linear regression model with one predictor measured with error,

$$y_i = \alpha_0 + \alpha_1 z_i + \beta_1 x_i + \varepsilon_i$$  \hspace{1cm} (4.2.2)$$

$$w_i = x_i + u_i$$  \hspace{1cm} (4.2.3)$$
In the above model, \( x_i \) is not observed or observed with substantial error and the covariate \( z_i \) is the error free predictor, hence \( i = 1, \ldots, n \) and \( \varepsilon_i \)'s are independently, identically distributed with mean 0 and constant but unknown variance \( \sigma^2_{ex} \). Besides this assume that the variance of \( u_i, \sigma^2_{wu} \) is known or estimated variance is available. Furthermore, in matrix notation, let,

\[
\begin{pmatrix}
X_1 \vspace{1pt} \\
X_2 \\
X_n
\end{pmatrix}, \quad \alpha = \begin{pmatrix}
\alpha_0 \\
\alpha_1
\end{pmatrix}, \text{ and } \alpha = \begin{pmatrix}
\alpha_0 \\
\alpha_1
\end{pmatrix}.
\]

Followed by Seber (1977) the least square estimate of the nuisance parameters \( \alpha = \begin{pmatrix}
\alpha_0 \\
\alpha_1
\end{pmatrix} \) are defined as,

\[
\hat{\alpha} = \begin{pmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1
\end{pmatrix} = \left[ X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right]^{-1} \left[ X_2'Y - X_2'X_1(X_1'X_1)^{-1}X_1'Y \right]
\]

Now define, \( y_i^* = y_i - \hat{\alpha}_0 - \hat{\alpha}_1z_i, \ i = 1, \ldots, n \).

It is convenient to work with normal equations; an approach based on using the normal equations as estimating equations for \( \beta_i \) requires somewhat weaker moment conditions and lead to simpler computations. If for some distribution \( F \) on pairs \((y_i^*, x_i)\), we have,

\[
\beta = \left( \int x'x \ dF \right)^{-1} \int x'y^* \ dF, \text{ and it implies, } \int x' \left( y_i^* - x\beta \right) dF = 0. \text{ Therefore, introduce the random variable,}
\]

\[
M_i = M_i(\beta) = x_i' \left( y_i^* - x_i\beta \right), \ i = 1, \ldots, n
\]
Subsequently, the $M_i$'s are independently, identically distributed by construction and $\beta_i = \beta$ if and only if $E(M_i) = 0$ and $M_i$ have a finite variance, that is, $E\|x_i'(v_i - x_i\beta)\|^2 < \infty$. Using the auxiliary variable $M_i$, calculate the estimated probability, $\hat{p}_i$, $i = 1, \ldots, n$. After that, find the empirical likelihood ratio statistic,

$$R_n(\beta_i) = -2 \sum_{i=1}^n \log (n\hat{p}_i)$$

The limiting distribution of $R_n(\beta_i)$ follows $\chi^2_{(1)}$ distribution and an asymptotic $100(1 - \alpha)$% confidence region for $\beta_i$ is given by,

$$S(\alpha) = \{ \beta : R_n(\beta_i) \geq \chi^2_{1}(\alpha) \}$$

Accordingly, $\chi^2{1}(\alpha)$ is upper $100$ $\alpha$% quantile of the $\chi^2_{1}$ distribution.

4.2.2 Application of the Method to Multiple Linear Regression Model: Example

To explain the application of QVF using empirical likelihood approach in estimating the regression parameters consider the sample of 55 managers of Iowa farmer cooperatives. This example was studied by Warren, White and Fuller (1974) predicting job performance of farm managers; in the original data there were 98 responses but here a sub sample of 55 individuals is selected. The data are given in the following table. This data will be employed for this section and for the next section.

Five measurements were made on each manager with the objective of studying managerial behavior: knowledge, value orientation, role satisfaction, past training and role performance.
The role behavior of a manager in farmer cooperatives is measured by his Role Performance and is linear function of the four variables

\[ X_1 : \text{Knowledge: knowledge of economic phases of management directed toward profit-making in a business and product knowledge;} \]

\[ X_2 : \text{Value Orientation: tendency to rationally evaluate means to an economic end;} \]

\[ X_3 : \text{Role Satisfaction: gratification obtained by the manager from performing the managerial role;} \]

\[ Z : \text{Past Training: amount of formal education.} \]
Table 4.1 Data from role performance study

<table>
<thead>
<tr>
<th>Observation</th>
<th>Knowledge, $X_1$</th>
<th>Value Orientation, $X_2$</th>
<th>Role Satisfaction, $X_3$</th>
<th>Past Training, $Z$</th>
<th>Role Performance, $Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1193</td>
<td>2656</td>
<td>2333</td>
<td>2</td>
<td>-0.054</td>
</tr>
<tr>
<td>2</td>
<td>1654</td>
<td>3300</td>
<td>2320</td>
<td>2</td>
<td>0.376</td>
</tr>
<tr>
<td>3</td>
<td>1193</td>
<td>2489</td>
<td>2737</td>
<td>2</td>
<td>0.072</td>
</tr>
<tr>
<td>4</td>
<td>1077</td>
<td>2478</td>
<td>2203</td>
<td>2</td>
<td>-0.15</td>
</tr>
<tr>
<td>5</td>
<td>1539</td>
<td>2822</td>
<td>2840</td>
<td>2</td>
<td>0.171</td>
</tr>
<tr>
<td>6</td>
<td>1385</td>
<td>3000</td>
<td>2373</td>
<td>2</td>
<td>0.042</td>
</tr>
<tr>
<td>7</td>
<td>1462</td>
<td>3111</td>
<td>2497</td>
<td>2.667</td>
<td>0.188</td>
</tr>
<tr>
<td>8</td>
<td>1385</td>
<td>2545</td>
<td>2617</td>
<td>2.167</td>
<td>-0.052</td>
</tr>
<tr>
<td>9</td>
<td>1539</td>
<td>2556</td>
<td>2997</td>
<td>2</td>
<td>0.31</td>
</tr>
<tr>
<td>10</td>
<td>1654</td>
<td>2945</td>
<td>2150</td>
<td>2.167</td>
<td>0.147</td>
</tr>
<tr>
<td>11</td>
<td>1462</td>
<td>2778</td>
<td>2227</td>
<td>1.833</td>
<td>0.005</td>
</tr>
<tr>
<td>12</td>
<td>1154</td>
<td>2545</td>
<td>2017</td>
<td>2.167</td>
<td>-0.088</td>
</tr>
<tr>
<td>13</td>
<td>1424</td>
<td>3611</td>
<td>2303</td>
<td>2.333</td>
<td>0.044</td>
</tr>
<tr>
<td>14</td>
<td>1116</td>
<td>2956</td>
<td>2517</td>
<td>2.333</td>
<td>-0.073</td>
</tr>
<tr>
<td>15</td>
<td>1270</td>
<td>2856</td>
<td>1770</td>
<td>2.333</td>
<td>0.224</td>
</tr>
<tr>
<td>16</td>
<td>1347</td>
<td>2956</td>
<td>2430</td>
<td>2</td>
<td>0.103</td>
</tr>
<tr>
<td>17</td>
<td>1116</td>
<td>2545</td>
<td>2043</td>
<td>2</td>
<td>-0.108</td>
</tr>
<tr>
<td>18</td>
<td>1077</td>
<td>3356</td>
<td>2410</td>
<td>2</td>
<td>-0.019</td>
</tr>
<tr>
<td>19</td>
<td>1423</td>
<td>3211</td>
<td>2150</td>
<td>2</td>
<td>-0.062</td>
</tr>
<tr>
<td>20</td>
<td>923</td>
<td>2556</td>
<td>2180</td>
<td>2</td>
<td>-0.239</td>
</tr>
<tr>
<td>21</td>
<td>1385</td>
<td>2589</td>
<td>2490</td>
<td>2</td>
<td>-0.159</td>
</tr>
<tr>
<td>22</td>
<td>1270</td>
<td>2900</td>
<td>1920</td>
<td>2.333</td>
<td>0.069</td>
</tr>
<tr>
<td>23</td>
<td>1116</td>
<td>2167</td>
<td>2663</td>
<td>1.333</td>
<td>-0.118</td>
</tr>
<tr>
<td>24</td>
<td>1346</td>
<td>2922</td>
<td>2520</td>
<td>2</td>
<td>0.083</td>
</tr>
<tr>
<td>25</td>
<td>846</td>
<td>1711</td>
<td>3150</td>
<td>1.5</td>
<td>-0.255</td>
</tr>
<tr>
<td>26</td>
<td>1077</td>
<td>2556</td>
<td>2297</td>
<td>2</td>
<td>-0.159</td>
</tr>
<tr>
<td>27</td>
<td>1231</td>
<td>3567</td>
<td>2307</td>
<td>2.167</td>
<td>0.014</td>
</tr>
<tr>
<td>Observation</td>
<td>Knowledge, $X_1$</td>
<td>Value Orientation, $X_2$</td>
<td>Role Satisfaction, $X_3$</td>
<td>Past Training, $Z$</td>
<td>Role Performance, $Y$</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>28</td>
<td>962</td>
<td>2689</td>
<td>2830</td>
<td>1.333</td>
<td>0.102</td>
</tr>
<tr>
<td>29</td>
<td>1500</td>
<td>2978</td>
<td>2737</td>
<td>2</td>
<td>0.109</td>
</tr>
<tr>
<td>30</td>
<td>1577</td>
<td>2945</td>
<td>3117</td>
<td>2.167</td>
<td>0.006</td>
</tr>
<tr>
<td>31</td>
<td>1885</td>
<td>3256</td>
<td>2647</td>
<td>2.667</td>
<td>0.334</td>
</tr>
<tr>
<td>32</td>
<td>1231</td>
<td>2956</td>
<td>2217</td>
<td>1.667</td>
<td>-0.076</td>
</tr>
<tr>
<td>33</td>
<td>1808</td>
<td>2811</td>
<td>2327</td>
<td>2</td>
<td>-0.043</td>
</tr>
<tr>
<td>34</td>
<td>1039</td>
<td>2733</td>
<td>2447</td>
<td>2</td>
<td>-0.126</td>
</tr>
<tr>
<td>35</td>
<td>1385</td>
<td>2400</td>
<td>2347</td>
<td>2.167</td>
<td>-0.056</td>
</tr>
<tr>
<td>36</td>
<td>1846</td>
<td>2944</td>
<td>2410</td>
<td>2</td>
<td>0.203</td>
</tr>
<tr>
<td>37</td>
<td>1731</td>
<td>3200</td>
<td>2277</td>
<td>2</td>
<td>0.023</td>
</tr>
<tr>
<td>38</td>
<td>1500</td>
<td>2911</td>
<td>2577</td>
<td>2</td>
<td>0.047</td>
</tr>
<tr>
<td>39</td>
<td>1231</td>
<td>3167</td>
<td>2507</td>
<td>2.333</td>
<td>0.011</td>
</tr>
<tr>
<td>40</td>
<td>1346</td>
<td>3322</td>
<td>2653</td>
<td>2.5</td>
<td>0.153</td>
</tr>
<tr>
<td>41</td>
<td>1347</td>
<td>2833</td>
<td>2587</td>
<td>2.667</td>
<td>0.1</td>
</tr>
<tr>
<td>42</td>
<td>1154</td>
<td>2967</td>
<td>3140</td>
<td>2</td>
<td>-0.089</td>
</tr>
<tr>
<td>43</td>
<td>923</td>
<td>2700</td>
<td>2557</td>
<td>1.833</td>
<td>0.007</td>
</tr>
<tr>
<td>44</td>
<td>1731</td>
<td>3033</td>
<td>2423</td>
<td>2</td>
<td>0.089</td>
</tr>
<tr>
<td>45</td>
<td>1808</td>
<td>2911</td>
<td>2793</td>
<td>2</td>
<td>0.182</td>
</tr>
<tr>
<td>46</td>
<td>1193</td>
<td>3311</td>
<td>2283</td>
<td>2.333</td>
<td>0.259</td>
</tr>
<tr>
<td>47</td>
<td>1308</td>
<td>2245</td>
<td>2210</td>
<td>2</td>
<td>0.007</td>
</tr>
<tr>
<td>48</td>
<td>1424</td>
<td>2422</td>
<td>2350</td>
<td>2</td>
<td>-0.015</td>
</tr>
<tr>
<td>49</td>
<td>1385</td>
<td>2744</td>
<td>2330</td>
<td>2</td>
<td>-0.023</td>
</tr>
<tr>
<td>50</td>
<td>1385</td>
<td>2956</td>
<td>2130</td>
<td>2</td>
<td>-0.15</td>
</tr>
<tr>
<td>51</td>
<td>1347</td>
<td>2933</td>
<td>2837</td>
<td>2.167</td>
<td>0.152</td>
</tr>
<tr>
<td>52</td>
<td>1539</td>
<td>3411</td>
<td>2600</td>
<td>2.267</td>
<td>0.377</td>
</tr>
<tr>
<td>53</td>
<td>1385</td>
<td>1856</td>
<td>2790</td>
<td>2</td>
<td>0.043</td>
</tr>
<tr>
<td>54</td>
<td>1654</td>
<td>3089</td>
<td>2500</td>
<td>2</td>
<td>0.184</td>
</tr>
<tr>
<td>55</td>
<td>1308</td>
<td>2967</td>
<td>2813</td>
<td>2.667</td>
<td>0.127</td>
</tr>
</tbody>
</table>
To measure Knowledge, the specialists in the relevant fields of economics, fertilizers and chemicals formulated sets of questions. The measure of rational Value Orientation to economic ends was a set of 30 items administered to the respondents by questionnaire on which the respondents were asked to indicate the strength of their agreement or disagreement. The respondents indicated the strength of satisfaction or dissatisfaction for each of 11 statements covering four areas of satisfaction:

i. Managerial role itself

ii. The position

iii. Rewards and,

iv. Performance of complementary role players.

The amount of past training was the total years of formal schooling divided by six. Role Performance was measured with a set of 24 questions covering the five functions of planning, organizing, controlling, coordinating and directing. The recorded verbal responses of managers of how they performed given tasks were scored by judges on a scale of 1 to 99 on the basis of performance leading to successful management. Responses to each question were randomly presented to judges and the raw score were transformed by obtaining the $Z$-value for areas of 0.01 to 0.99 from a cumulative normal distribution (a raw score of 40 received transformed score of $-0.253$). For each question, the mean of transformed scores of judges was calculated. The first three independent variables were obtained on the basis of qualitative and quantitative answers to identical questionnaires distributed to the selected random sample of managers in farmers’ cooperatives. These variables are assumed to be measured with error. The remaining independent variable, Past Training measured on the basis of a single quantitative answer assumed error free. Furthermore, the errors in the measurement of these variables are mutually independent.
Using replicated determinations on the same individuals; measurement error variances were estimated to be 0.0037, 0.0203, 0.0438, and 0.0180 for role performance, knowledge, value orientation, and role satisfaction, respectively. Each error variance estimate is based on 97 degrees of freedom and it is assumed that the error variance is diagonal.

![Knowledge VS Role Performance](image)

**Figure 4.1 Scatter plot of Knowledge and Role Performance**

Before performing any statistical analysis, we use scatter plots to explore the data visually. First we examine the scatter plot of Knowledge versus Role Performance for any distinct patterns and for any outliers, which are points far away from all other points. The role performance of the managers on Iowa farmer cooperatives is significantly correlated with their knowledge (at the 1% level of significance level with p-value< 0.00001); the correlation
coefficient in this case is 0.588. The Figure 4.1 also gives an impression that a straight line may describe the relationship between knowledge and role performance.

![Value Orientation VS Role Performance](image)

Figure 4.2 Scatter plot of Value Orientation and Role Performance

The pattern made by the points (Value Orientation, Role Performance) plotted on the scatter diagram, Figure 4.2 suggests a basic nature and the strength of the relationship. The plot shows that the managers of the Iowa farmer cooperatives with large value orientations also provide better role performance. These impressions suggest that a straight line crossing the y-axis below the origin may describe the relationship between two variables; actually there is a statistically significant (at the 15 level of significance with p-value < 0.001) relationship between the value orientation and role performance. The correlation coefficient was obtained as 0.495.
Figure 4.3 Scatter plot of Role Satisfaction and Role Performance

The scatter plot in Figure 4.3 shows no distinct pattern between the role satisfaction of the managers and their role performance; consequently, the correlation coefficient (at the 1% significance level with p-value=0.2973) becomes 0.143 between these two variables.
The plot in Figure 4.4 suggests a positive relationship between the past training of the managers and their performance; the correlation coefficient is 0.408 (at the 1% significance level with p-value < 0.002).

Now for the model (4.2.2), adopt only the predictors, knowledge and past training and the response, role performance. As a result, the model becomes;

\[
(y_i) = \alpha_0 + \alpha_1 (Past Training)_i + \beta_1 (Knowledge)_i + \epsilon_i \tag{4.2.4}
\]

Since “knowledge” is obtained on the basis of responses to several questions, so there is measurement error with it. Now, define \( y_i \) is the role performance of the \( ith \) manager, \( x_i \) is the
measure of the knowledge of economic phases of management of the \( i \)th individual and \( z_i \) is his past training which is error free covariate. In matrix notation, let,

\[
X_1 = \begin{pmatrix} x_1 \\ \vdots \\ x_{55} \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_{55} \end{pmatrix}, \text{ and } \alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}.
\]

Following Sever (1977) the least squared estimate of the regression parameters are obtained as;

\[
\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{pmatrix} = \begin{bmatrix} X_2'X_2 - X_2'X_1 (X_1'X_1)^{-1} X_1'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_2'Y - X_2'X_1 (X_1'X_1)^{-1} X_1'Y \end{bmatrix}
\]

and

\[
\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{pmatrix} = \begin{pmatrix} -0.664 \\ 0.14 \end{pmatrix}
\]

at this time, considering the quasi likelihood variance function for the model (4.2.4), the auxiliary variable is attained,

\[
M_i = M_i(\beta) = x_i' (y_i^* - x_i \beta),
\]

hence,

\[
y_i^* = (Role \ Perfoemence)_i - \hat{\alpha}_0 - \hat{\alpha}_1(Past \ Training)_i, \ i = 1, \ldots, 55.
\]

By following the empirical likelihood method, the estimate of the regression coefficient of the error prone variable is obtained. The listed Table 4.2 shows the comparison of the estimators obtained from naïve, regression calibration, SIMEX, and QVF and EL method for the linear regression model with single error prone covariate.
Table 4.2 Comparison of Estimators: Naïve, Regression Calibration, SIMEX, and QVF and EL Method for the Linear Regression Model with Single Variable Measured with Error

<table>
<thead>
<tr>
<th>Method of estimating regression parameters</th>
<th>Estimate of $\hat{\beta}_1$, $\hat{\beta}_i$</th>
<th>Confidence Interval for $\hat{\beta}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Method</td>
<td>0.308</td>
<td>(0.173, 0.439)</td>
</tr>
<tr>
<td>Regression Calibration</td>
<td>0.842</td>
<td>(0.185, 1.499)</td>
</tr>
<tr>
<td>SIMEX</td>
<td>0.359</td>
<td>(0.195, 0.523)</td>
</tr>
<tr>
<td>QVF and EL Method</td>
<td>-</td>
<td>(-0.314, 0.382)</td>
</tr>
</tbody>
</table>

The above Table 4.2 illustrates the performance of the proposed method for the linear regression model, wherein only the single predictor, Knowledge is subject to error and the effect of measurement error cannot be determined. Here we assess the quality of the various methods in terms of the performance of the parameter estimates, $\hat{\beta}_i$ regarding confidence intervals. In the proposed estimation approach, the estimate of the regression parameter of the error prone predictor is obtained by considering QVF measurement error model and then applying empirical likelihood method to solve the mismeasured problem. For this real life example, the value of estimated slopes are almost close in naïve and SIMEX approach and the most popular method, regression calibration gives the largest confidence interval while the naïve provides the shortest one. The proposed method, wherein the confidence interval is calculated considering mean and variance function of the measurement error model and employing the EL method on those functions; presents the shorter confidence interval than the most popular method regression calibration ($0.696 < 1.314$) for the regression parameter of the error prone variable in this example.

4.3 Multiple Regression: Several Covariates Measured with Error

In the example of section (4.2.2), the estimate of a single error prone covariate was obtained; in this section we will consider all three error prone covariates, Knowledge, Value...
Orientation and Role Satisfaction to show the application of the QVF and empirical likelihood approach. Therefore the model considering all the measurement is,

\[
\begin{pmatrix}
\text{Role Performance}_i \\
\text{Training}_i
\end{pmatrix} = \alpha_0 + \alpha_1 \begin{pmatrix}
\text{Past Knowledge}_i \\
\text{Value Orientation}_i \\
\text{Role Satisfaction}_i
\end{pmatrix} + \beta_1 \text{Role Performance}_i + \beta_2 \text{Training}_i + \beta_3 \text{Value Orientation}_i + \beta_4 \text{Role Satisfaction}_i + \varepsilon_i
\]

(4.3.1)

That is,

\[
y_i = \alpha_0 + \alpha_1 z_i + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i
\]

(4.3.2)

Consequently, \( y_i \) is the role performance of the \( ith \) manager, \( x_{1i} \) is the measure of the knowledge of economic phases of management, \( x_{2i} \) is the value orientation and \( x_{3i} \) is the role satisfaction of the \( ith \) individual and \( z_i \) is his past training which is error free covariate.

In matrix notation, let,

\[
Y = \begin{pmatrix}
y_1 \\
\vdots \\
y_{55}
\end{pmatrix}_{55 \times 1}, \quad X_1 = \begin{pmatrix}
x_{11} & x_{21} & x_{31} \\
\vdots & \ddots & \vdots \\
x_{155} & x_{255} & x_{355}
\end{pmatrix}_{55 \times 3}, \quad X_2 = \begin{pmatrix}
1 \\
\vdots \\
1
\end{pmatrix}_{55 \times 1}, 
\]

\[
\alpha = \begin{pmatrix}
\alpha_0 \\
\alpha_1
\end{pmatrix}_{2 \times 1}, \quad \beta = \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3
\end{pmatrix}_{3 \times 1}.
\]

The least squared estimate of the regression coefficients from Seber (1977) becomes,

\[
\hat{\alpha} = \begin{pmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1
\end{pmatrix} = \left[ X_2'X_2 - X_2'X_1 \left( X_1'X_1 \right)^{-1} X_1'X_2 \right]^{-1} \left[ X_2'Y - X_2'X_1 \left( X_1'X_1 \right)^{-1} X_1'Y \right]
\]

and analyzing the data the estimates are obtained as,
Now considering mean and variance function as discussed in section (4.1.2), the auxiliary variable for this example is obtained as,

\[ M_i(\beta) = (y_i - \bar{x}_i \beta) \bar{x}_i \]  

(4.3.3)

Here, \( \bar{x}_i = (x_{i1}, x_{i2}, x_{i3}) \) and \( y_i^* = y_i - \hat{\alpha}_0 - \hat{\alpha}_1 z_i, \ i = 1, \ldots, 55 \). Using the value of the auxiliary variable for each individual in the data set define the estimated probability, \( \hat{p}_i \) (see, equation 4.1.16) of including that individual in that particular sample. Based on this estimated probability, obtain the empirical likelihood ratio statistic for the inference of \( \beta \), which is

\[ R_n(\beta) = -2 \sum_{i=1}^{n} \log (n \hat{p}_i) \]

Consequently, the limiting distribution of \( R_n(\beta) \) follows \( \chi^2_3 \) distribution and an asymptotic 100 \((1 - \alpha)\%\) confidence region for \( \beta \) is given by,

\[ S(\alpha) = \{ \beta : R_n(\beta) \geq \chi^2_3(\alpha) \} \]

at this moment, \( \chi^2_3(\alpha) \) is upper 100 \( \alpha \%) quantile of the \( \chi^2_3 \) distribution.

The following Table 4.3 shows the comparison of estimators; like Naïve, SIMEX, and QVF and EL Method for the multiple linear regression model with several variables measured with error.
Table 4.3 Comparison of Estimators: Naïve, SIMEX and QVF and EL Method for the Multiple Linear Regression Model with Several Variables Measured with Error

<table>
<thead>
<tr>
<th>Regression Parameters</th>
<th>Naïve Method</th>
<th>SIMEX</th>
<th>QVF and EL Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.2526</td>
<td>0.2867</td>
<td>-</td>
</tr>
<tr>
<td>CI for $\hat{\beta}_1$</td>
<td>(0.115, 0.391)</td>
<td>(0.142, 0.432)</td>
<td>(-0.314, 0.382)</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.1111</td>
<td>0.1438</td>
<td>-</td>
</tr>
<tr>
<td>CI for $\hat{\beta}_2$</td>
<td>(0.013, 0.209)</td>
<td>(0.037, 0.251)</td>
<td>(-0.509, 0.761)</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>0.0967</td>
<td>0.1295</td>
<td>-</td>
</tr>
<tr>
<td>CI for $\hat{\beta}_3$</td>
<td>(-0.009, 0.202)</td>
<td>(-0.017, 0.276)</td>
<td>(-0.539, 0.632)</td>
</tr>
</tbody>
</table>

4.4 Comparisons Among Different Techniques Reducing Measurement Error Bias

Several methods have been developed to adjust for the bias resulting when one or more regression covariates are measured with error. Regression calibration is one method, which is widely used and easy to implement. It has been discussed and applied in a variety of contexts. The inference based on this method seeks to estimate the regression models with measurement error in explanatory variables by replacing the mismeasured variable by its conditional expectation in an estimation procedure that would have been used if the true variable were available. Since the covariate is replaced by its estimated expectation, it becomes an imprecise measurement, subject itself to the problem of regression with measurement error. Regression calibration estimator shows an improvement compared to the naïve regression of the response and observed but not true value of the predictor, because the estimated conditional expectation should be closer to the true conditional expectation than the observed value is to true value. But frequently, the resulting estimators become seriously biased if the estimated expectation is not sufficiently close to the actual value. Most of the cases regression calibration ignores the effect
of the uncertainty on the estimation of, either by assuming that is known, or assuming that it is consistent and basing inference on the asymptotic distribution of the regression calibration estimator. Regression calibration approach cannot be applied to any measurement error problem; to employ this approach it is assume that the measurement error is non differential. According to Carroll (Chapter 4, 2006) it is simply a poor person’s imputation methodology. Furthermore, the distribution of the error prone covariate is not allowed to depend on observed covariates that are measured without error in this method. Another limitation is that the methods work well for one variable subject to measurement error, and they do not easily generalize to multiple correlated variables. Finally, the computational burden seems prohibitive for data sets with very large number of observations and multiple covariates.

The score function method is another well-liked technique for reducing the measurement error bias. In addition, this method is extensively used in the linear models measurement error settings but is not a viable option for nonlinear measurement error models (see Carroll, 2006). Moreover, the functional estimator in most nonlinear models is both extremely difficult to compute and not even consistent or valid. This method does not make any assumption about the error prone covariates which makes the computation easier and leads to valid estimation and inference. Although the mathematical notation of conditional and corrected scores are more complex than that of regression calibration and SIMEX, the formula are simple to program and implement.

4.5 Simulation Study

We conducted a simulation study to illustrate the performance of the proposed method for a multiple linear regression model, wherein two predictors are subject to error and a single covariate is measured without error. To examine the method we consider the following model:
Here instead of observing $X$, one observes $W$, therefore the primary data set contains $(y, z, w)$; where $e_i$ has a standard normal distribution. To set the mean and variance function, we assume that the replications or instrumental data are available to obtain an estimate of $\hat{x}_i$. For this study, a sample of size 50 of the random error is generated from a standard normal distribution. Then the response variable \( \{ y_i, \quad i = 1, \ldots, n \}; \ n = 50 \) is generated by considering $\alpha_0 = 1, \alpha_1 = 2, \beta_0 = 2, \beta_2 = 3$; and it also follows normal distribution. Therefore, the conditional distribution of $y_i$ given $w_i$ is also normal. The following Table 4.4 provides an impression of the proposed method. We generate a sample of size, $n=50$ for $m = 100, 200, 500, 1000, \text{ and } 2000 \text{ times}$; then we find the 95\% confidence intervals for the coefficients of the error prone covariates, and obtain the average length of the intervals and calculate the percentage of times the intervals capture the regression parameters of the misspecified covariates.
Table 4.4 Simulation Output

<table>
<thead>
<tr>
<th>Simulation size, $m$</th>
<th>Average Length of CI for $\hat{\beta}_0$</th>
<th>Average Length of CI for $\hat{\beta}_1$</th>
<th>% of Times the CI Captures the parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>9.42</td>
<td>15.23</td>
<td>97%</td>
</tr>
<tr>
<td>200</td>
<td>9.67</td>
<td>15.59</td>
<td>100%</td>
</tr>
<tr>
<td>500</td>
<td>9.65</td>
<td>15.60</td>
<td>99.2%</td>
</tr>
<tr>
<td>1,000</td>
<td>9.73</td>
<td>15.66</td>
<td>99.2%</td>
</tr>
<tr>
<td>2,000</td>
<td>9.61</td>
<td>15.38</td>
<td>99.3%</td>
</tr>
</tbody>
</table>

In all empirical or experimental sciences, it is a standard approach to present results, additionally to point estimates, in form of confidence intervals of the parameters of interest. In many problems of statistical inference, the experimenter is interested in constructing a family of sets that contain the true (unknown) parameter value with a specified (high) probability. The length of a confidence interval characterizes the accuracy of the whole findings. Clearly, the larger the width of the interval, the better is the chance of trapping a true parameter value. Confidence intervals provide a range about the observed estimates of the regression parameters. This range is constructed in such a way that we know how likely it is to capture the true but unknown regression coefficients. Thus the formal definition of a confidence interval is: a range of values for a variable of interest [in our case, the estimated regression parameters of the error prone covariates] constructed so that this range has a specified probability of including the true value of the variable. For this study, the probability of capturing the true, unknown parameter values are quite large (97%-100%) and such intervals are desirable since they are more informative. Again the average widths of the confidence intervals are reasonably large and also consistent for different simulated samples, which convey some useful information. If the confidence interval is narrow, capturing only a small range of true values, we can be quite
confident that any estimated values far from this range have been ruled out by the study. But the proposed method provides logically large width with high inclusion probability. The proposed technique for finding confidence intervals for regression coefficients of the covariates subject to error are computationally easy to employ to any measurement error models for correcting bias due to it.

4.6 Estimation of True Values

Consider the classical linear regression model with one independent variable defined as,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ i = 1, \ldots, n \quad (4.6.1)$$

But one is unable to observe $x_i$ directly; instead of observing $x_i$, one observes $w_i$, where $w_i$ is defined as,

$$w_i = x_i + u_i \quad (4.6.2)$$

Hence, $u_i$ is a random variable with mean 0 and variance $\sigma^2_u$; and

$$(x_i, \epsilon_i, u_i) \sim NI\left((\mu_x, 0, 0)', \text{diag}(\sigma_{xx}, \sigma_{\epsilon \epsilon}, \sigma_{uu})\right) \quad (4.6.3)$$

In this section an estimator of $x$-value is constructed that generates the vector $V_i = (Y_i, W_i)$. In constructing an estimator of $x_i$, the unknown $x_i$ can be treated as fixed or random and these two assumptions lead to different procedures.

By treating $x_i$ as fixed unknown constants; let the error assumptions of the model (4.6.1) hold and let $\beta_0, \beta_1, \sigma_{\epsilon \epsilon}$ and $\sigma_{uu}$ be known and define,

$$y_i - \beta_0 = \beta_1 x_i + \epsilon_i \quad (4.6.4)$$
Equation (4.6.4) is given in the form of the classical regression model, where \( x_i \) is the unknown parameter to be estimated. Therefore, if we treat \( x_i \) as a fixed unknown constant, the best linear unbiased estimator of \( x_i \) is given by the generalized least square estimator;

\[
\hat{x}_i = \left( \beta_1 \Sigma^{-1} \beta_1 \right)^{1\prime} \left( y_i - \beta_0 \right)^\prime
\]  

(4.6.5)

Where \( \Sigma \) is the covariance matrix \((\epsilon_i \quad u_i)\); if \( \Sigma = diag(\sigma_{\epsilon \epsilon} \quad \sigma_{u u}) \), the expression (4.4.5) reduces to

\[
\hat{x}_i = \left( \beta_1^2 \sigma_{u u} + \sigma_{\epsilon \epsilon} \right)^{1\prime} \left( \sigma_{u u} (y_i - \beta_0) \beta_1 + \sigma_{\epsilon \epsilon} w_i \right)
\]  

(4.6.6)

The variance of the estimator (4.4.5) becomes,

\[
Var(\hat{x}_i - x_i) = \left( \beta_1 \Sigma^{-1} \beta_1 \right)^{1\prime}
\]  

(4.6.7)

And for the expression (4.4.6) when \( \Sigma \) is diagonal, the variance of the estimator becomes,

\[
Var(\hat{x}_i - x_i) = \left[ \beta_1^2 \sigma_{\epsilon \epsilon}^{-1} + \sigma_{u u}^{-1} \right]
\]  

(4.6.8)

If we consider the yield of corn and available soil nitrogen where the fields were a random sample of farmers’ fields in the state of Iowa, the true value of the soil nitrogen can be treated as random variables. Under the assumption of random \( x_i \), the population covariance matrix of \((Y_i, W_i, X_i)\) is,

\[
\Sigma_{pop} = \begin{bmatrix}
\beta_1^2 \sigma_{\epsilon \epsilon} + \sigma_{\epsilon \epsilon} & \beta_i \sigma_{\epsilon \epsilon} + \sigma_{\epsilon \epsilon} & \beta_i \sigma_{\epsilon \epsilon} \\
\beta_i \sigma_{\epsilon \epsilon} + \sigma_{\epsilon \epsilon} & \sigma_{\epsilon \epsilon} + \sigma_{\epsilon \epsilon} & \sigma_{\epsilon \epsilon} \\
\beta_i \sigma_{\epsilon \epsilon} & \sigma_{\epsilon \epsilon} & \sigma_{\epsilon \epsilon}
\end{bmatrix}
\]  

(4.6.9)

Therefore, under the assumption of normality, the expected value of \( x_i \) is given \( V_i = (Y_i, W_i) \) becomes,
\[
E\{X_i \mid Y_i, W_i\} = \mu_x + \kappa_1 (Y_i - \mu_y) + \kappa_2 (w_i - \mu_w)
\]
\[
= \kappa_0 + \kappa_1 Y_i + \kappa_2 W_i
\]  
(4.6.10)

Hence, \( \kappa_0 = (1 - \kappa_2) \mu_x - \kappa_1 \mu_y \), \( \mu_y = \beta_0 + \beta_1 \mu_x \) and \( (\kappa_1, \kappa_2)' = \sum_w^{-1} (\beta_1 \sigma_{xx}, \sigma_{xx})' \).

Using the fact that,

\[
\sum_{iv} = (\beta_1 \ 1)' \sigma_{xx} (\beta_1 \ 1) + \text{Cov}(\epsilon, u)
\]  
(4.6.11)

Therefore the expected value becomes,

\[
E\{X_i \mid Y_i, W_i\} = W_i - (V_i - \mu_v) \sum_{vw}^{-1} \sum_{vu}
\]  
(4.6.11)

Where, \( \sum_{vu} = \sum_{cu} = (\sigma_{xu}, \sigma_{uw})' \) and \( e_i = (\epsilon_i, u_i) \), the conditional variance of \( x_i \) is,

\[
\text{Var} (X_i \mid Y_i, W_i) = \sigma_{xx} + (\kappa_1 \ \kappa_2)' (\beta_1 \sigma_{xx} \ \sigma_{xx})'
\]  
(4.6.12)

In both situations, the estimator is obtained by subtracting a predictor of \( u_i \) from \( w_i \).
CHAPTER 5
SUMMARY AND CONCLUSION

Ignoring measurement error may have caused bias in the estimation of regression parameters. In recent years, much work has been done in the context of measurement error. Several approaches have been developed either to make explicit correction for the bias in parameter estimation or to give alternative estimators under certain assumptions. We suggest a simple computational estimation procedure that gives confidence intervals for error-prone regression parameters. This approach is based on quasi-likelihood and variance function (QVF) and empirical likelihood method.

Each chapter of this dissertation offers specific ideas or recommendations concerning ways in which researchers can guard against making erroneous decisions about the parameters in regression model because of measurement error. Chapter 1 develops to provide ideas of measurement error and its effect on the linear regression model. It also presents the existing models for measurement error and ideas for correcting bias. The purpose of Chapter 2 is to demonstrate the description of likelihood and quasi-likelihood method for measurement error model with their necessity in the statistical analysis with an example. This chapter displays the application of these approaches for both univariate and multivariate cases. Chapter 3 presents the different existing methods to deal with measurement error problem in regression models. The most popular applicable approaches to measurement error analyses called regression calibration and simulation extrapolation are discussed in chapter 3. The score function method, another widely applicable method for eliminating or reducing measurement error bias, is also presented here. In this chapter, we also apply these approaches to a real-life example and compare the results. We propose a simple computationally easy estimation procedure for the parameters of
error prone covariates in regression models in Chapter 4. This chapter also presents comparisons of the proposed estimation approach with exiting methods. Applications of this approach to the real life data are also shown. A simulation study is also carried out to show the performance of the new approach. It is showed that the capturing probabilities of the unknown parameters are very large in the proposed approach, which is highly desirable to set the confidence intervals of unknown regression coefficients of the error prone covariates.

In this dissertation, we consider simple and multiple linear regression models with one or more continuous predictors subject to measurement error. However, we can extend this work for discrete error-prone covariates. Additional work can be done when both response and predictors are measured with error. We can further illustrate the application of the proposed approach to nonlinear models with misspecified predictors.
REFERENCES


*Biometrika*, **72**, 583-592.


APPENDICES

A.1 Likelihood Method

The likelihood method is a commonly used method for obtaining an estimate of an unknown parameter of an assumed population distribution. The likelihood of a data set depends upon the parameter(s) of the distribution or probability density function from which the observations have been taken. In cases where one or more of these parameters are unknown, a smart choice as an estimate would be the value that maximizes the likelihood. This is the maximum likelihood estimate (MLE). Expressions for maximum likelihood estimates are frequently obtained by maximizing the natural logarithm of the likelihood rather than the likelihood itself. Sir Ronald Fisher introduced the method in 1912.

The purpose of the likelihood function is to convey information about unknown quantities. The “information” is incomplete, and the function will express the degree of incompleteness. Unknown quantities in statistical problems may be fixed parameters, with the associated estimation problem, or unobserved random values; in a real prediction problem the two unknown can be easily mixed. Suppose for a moment that $X$ can take on only a countable set of values $x_1, x_2, \ldots$, with $Pr_\theta(x) = Pr_\theta(X = x)$, and that one wish to determine the correct value of $\theta$, that is, the value that produced the observed $x$. This suggests considering for each possible $\theta$ how probable the observed $x$ would be if $\theta$ were the true value. The higher this probability, the more one is attracted to the explanation that the $\theta$ in question produced $x$, and the more likely the value of $\theta$ appears. Therefore, the expression $Pr_\theta(x)$ considered for fixed $x$ as a function of $\theta$ has been called the likelihood of $\theta$. Let denote the probability $Pr_\theta(x)$ by $L(\theta; x_1, \ldots, x_n)$; then the definition becomes:
Assuming a statistical model parameterized by a fixed and unknown $\theta$, the likelihood $L(\theta; x_1, \ldots, x_n)$ is the probability of the observed data $x$ considered as a function of $\theta$.

The generic data $x$ include any set of observations from an experiment of any complexity: a range of values rather than exact measurements, a vector of values, a matrix, an array of matrices, a time series or a 2D image. The generic parameter $\theta$ can also be as complex as the model requires.

### A.1.1 Maximum Likelihood Estimation

Maximum likelihood estimation is reasonably easy to implement, efficient, and the basis of readily available inferential methods, such as standard errors of Fisher information and likelihood ratio tests. Also many other common estimators are closely related to maximum likelihood estimators, for example, the least squares estimator, quasi likelihood estimators, etc.

The likelihood for the discrete case becomes:

$$ L(\theta; x_1, \ldots, x_n) = L(\theta) = \prod_{i=1}^{n} p(x_i; \theta) $$

and for continuous data,

$$ L(\theta; x_1, \ldots, x_n) = L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \, dx_i $$

where $X_i$ has been observed to lie in a small set of volume $dx_i$ near the value $x_i$. Define $L(\theta; X_1, \ldots, X_n)$ for a random likelihood taking the above when $X_i = x_i$. Now let $\theta_0$ be the true value of $\theta$. The method of maximum likelihood estimates $\theta_0$ by finding the value $\hat{\theta}$ that maximize $L(\cdot)$:

$$ \hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; X_1, \ldots, X_n). $$
A maximizer $\hat{\theta}$ is called a maximum likelihood estimate (MLE) of $\theta_0$. It gives the best explanation of the data by making that data most probable. It is easier to work with the log likelihood function defined as

$$l(\theta) = \log L(\theta; x_1, \ldots, x_n) = c + \sum_{i=1}^{n} \log \left( f(x_i; \theta) \right)$$

where $c$ depends on the $dx_i$ but not on $\theta$. The most widely applied likelihood, $\hat{\theta}$ is found by solving the score equation,

$$\frac{\partial l(\theta)}{\partial \theta} = 0$$

or it may be written in terms of estimating equations,

$$\sum_{i=1}^{n} M(x_i; \theta) = 0 \quad (A.1.1)$$

and the function defined as,

$$M(x_i; \theta) = \frac{\partial}{\partial \theta} \frac{f(x; \theta)}{f(x; \theta)}$$

The vector of observations, including response, covariates, surrogates, etc., is denoted by $(\hat{Y}_i, Z_i)$ for $i = 1, \ldots, n$, where $Z_i$ is the covariates measured without error and $\hat{Y}_i$ collects all other variables in one vector and the vector $\Theta$ denotes the unknown parameters. The data set $(\hat{Y}_i, Z_i), i = 1, \ldots, n,$ is the aggregation of all data sets, primary and external, including replication and validation data. Hence $\hat{Y}_i$ depends on $i$, that is whether $i^{th}$ case is the validation case, a replication case etc., and $\hat{Y}_i$ are assumed independent, with density of $\hat{Y}_i$ depending on
both \( Z_i \) and the type of data set the \( i^{th} \) case came from and denoted by \( f_i(\tilde{y}; \Theta) \). Assume that \( f_i \) has two continuous derivatives with respect to \( \Theta \), then the likelihood becomes,

\[
L(\Theta) = \sum_{i=1}^{n} \log \left\{ f_i(\tilde{y}_i | \Theta) \right\}
\]

The maximum likelihood estimator (MLE) denoted by \( \hat{\Theta} \) maximizes the \( L(\Theta) \). Under some regularity conditions, the MLE has a simple asymptotic distribution. Now the likelihood score or the score function is obtained as

\[
S_i(y | \Theta) = \frac{\partial}{\partial \Theta} \log \left\{ f_i(y | \Theta) \right\}
\]

The Fisher information matrix, or expected information becomes,

\[
I_n(\Theta) = -\sum_{i=1}^{n} E \left\{ \frac{\partial}{\partial \Theta'} S_i(\tilde{y}_i | \Theta) \right\}
\]

(A.1.2)

In other words,

\[
I_n(\Theta) = \sum_{i=1}^{n} E \left\{ S_i(\tilde{y}_i | \Theta) \ S'_i(\tilde{y}_i | \Theta) \right\}
\]

(A.1.3)

In large samples, the MLE is approximately normally distributed with mean \( \Theta \) and covariance matrix \( I^{-1}_n(\Theta) \), whose entries converge to zero as \( n \to \infty \). The observed Fishers information matrix, an empirical version of (A.1.2) is obtained as,

\[
\hat{I}_n(\Theta) = -\sum_{i=1}^{n} E \left\{ \frac{\partial}{\partial \Theta'} S_i(\tilde{y}_i | \hat{\Theta}) \right\}
\]

(A.1.4)

The empirical version of (A.1.3) is,

\[
\hat{D}_n(\Theta) = \sum_{i=1}^{n} \left\{ S_i(\tilde{y}_i | \hat{\Theta}) \ S'_i(\tilde{y}_i | \hat{\Theta}) \right\}
\]

(A.1.5)

To estimate the \( I^{-1}_n(\Theta) \) use the sandwich formula, \( \hat{I}_n^{-1}(\Theta) \hat{D}_n^{-1}(\Theta) \hat{I}_n^{-1}(\Theta) \).
Maximum likelihood estimation is also widely used to construct estimates of the quantities that one might not know how to estimate.

### A.1.2 Likelihood Ratio and Confidence Intervals

For a random sample \( X_1, \ldots, X_n \) from a density \( f(x; \theta), \quad \theta \in \Omega \), and let \( \psi(\cdot) \) be a function of \( \theta \) and \( L(\theta; x_1, \ldots, x_n) \) be the likelihood function for the sample \( X_1, \ldots, X_n \) with joint density \( f_{x_1,\ldots,x_n}(x_1, \ldots, x_n; \theta) \) where \( \theta \in \Omega \). Suppose one wish to test \( H_0 : \psi(\theta) = 0 \) against the general alternative that \( H_1 : \psi(\theta) \neq 0 \). Then the likelihood ratio is defined as,

\[
\lambda = \lambda(x_1, \ldots, x_n; \theta) = \frac{L(\theta)}{L(\hat{\theta})}
\]  

(A.1.6)

The likelihood ratio \( \lambda \) is a function of \( x_1, \ldots, x_n \). Therefore, \( \lambda(X_1, \ldots, X_n; \theta) \) is a random variable and this statistic does not depend on unknown parameters. The values of the statistic \( \lambda \) are used to formulate a test \( H_0 \) versus \( H_1 \) by using the likelihood ratio principle, which states that \( H_0 \) is to be rejected if and only if \( \lambda \leq \lambda_0 \), where \( \lambda_0 \) is some fixed constant satisfying \( 0 \leq \lambda \leq 1 \). This likelihood ratio test makes good intuitive sense since \( \lambda \) will tend to be small when \( H_0 \) is not true, since then the denominator of \( \lambda \) tends to be larger than the numerator. The drawback of this test is that it is sometimes difficult to find \( \sup \{ L(\theta; x_1, \ldots, x_n) \} \); another is that it can be difficult to find the distribution of \( \lambda(X_1, \ldots, X_n; \theta) \) which is required to evaluate the power of the test.

In order to separate the reasonable values of \( \theta \) from the unreasonable one is to order them by \( \lambda \) and consider \( C = \{ \theta | \lambda > \lambda_0 \} \) fitting the data better than the other values of \( \theta \).
The way to pick $r_0$ is to aim for a given probability that $\theta_0 \in C$. Wilks’s theorem shows that

$$-2\log \lambda(\theta_0) \to \chi^2_{(1)}$$

distribution as $n \to \infty$. When it holds, use $C^{1-\alpha} = \{\theta \mid \lambda(\theta) \geq r_0\}$ as an approximate $1-\alpha$ confidence region for $\theta_0$ with $r_0 = \exp\left(-\frac{1}{2} \chi^2_{(1)}{1-\alpha}\right)$. To keep the calculation simple, consider the idea that for large $n$ the logarithm of likelihood ratio is very nearly quadratic around $\hat{\theta}$ under regularity conditions. A Taylor approximation shows,

$$l(\theta) = l(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^\top \hat{I}(\theta - \hat{\theta})$$

near $\theta$; where $l(\theta) = \log \lambda(\theta)$ and $\hat{I}$ is the Hessian of $l$ at $\hat{\theta}$.

The confidence region becomes,

$$\left\{ \theta \mid (\theta - \hat{\theta})^\top \hat{I}(\theta - \hat{\theta}) \leq \chi^2_{(1)}{1-\alpha}\right\}$$

(A.1.7)

Under standard assumptions (A.1.7) has asymptotic probability $1-\alpha$ of containing $\theta_0$.

A.2 The Bootstrap Idea

The bootstrap is widely used tool for analyzing the sampling variability of complex statistical methods. The basic idea is to create simulated data sets, whose distribution is equal to an estimate of the probability distribution of the actual data. Any statistical method that is applied to the actual data can also be applied to the bootstrap data sets.

Thus the empirical distribution of an estimator or test statistic across the bootstrap data sets can be used to estimate the actual sampling distribution of that statistic.

Suppose $\hat{\Theta}$ is obtained by applying some estimator to the actual data, and let $\hat{\Theta}^{(m)}$ is obtained by applying the same estimator to the $m^{th}$ bootstrap data set, $m = 1, \ldots, M$, where $M$ is the number of bootstrap data sets and let $\overline{\Theta} = \frac{1}{M} \sum_{m=1}^{M} \hat{\Theta}^{(m)}$, then the covariance matrix becomes,
\[
\text{vár}(\hat{\Theta}) = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\Theta}^{(m)} - \bar{\Theta}) (\hat{\Theta}^{(m)} - \bar{\Theta})^T
\]  
\hspace{1cm} (A.2.1)

There are many ways to estimate the probability distribution of the data, it is not always obvious which way is most appropriate. Bootstrap standard error can easily be obtained from (A.2.1) and plugged into the following expression to get the normal theory confidence intervals;

\[
\hat{\Theta} \pm \phi^{-1}\left(1 - \frac{\alpha}{2}\right)SE(\hat{\Theta})
\]  
\hspace{1cm} (A.2.2)

Moreover, the estimated standard error is defined as, 

\[SE(\hat{\Theta}) = \sqrt{\text{var}(\hat{\Theta})}.
\]

\textit{A.2.1 The Bootstrapping Regression Models}

Consider the QVF model;

\[Y_i = \mu (Z, \beta) + \sigma \gamma (Z, \beta, \theta) \varepsilon_i\]  
\hspace{1cm} (A.2.3)

Hence \(\varepsilon_i\)'s are independently and identically distributed. To estimate the sampling distribution of the QVF estimators, resampling from the set of pairs \(\{(Y_i, Z_i)\}_{i=1}^{n}\) can form bootstrap data sets. Then the \(i^{th}\) residual becomes,

\[e_i = \frac{Y_i - \mu(Z_i, \hat{\beta})}{\hat{\sigma} \gamma(Z_i, \hat{\beta}, \hat{\theta})} - \bar{e}\]  
\hspace{1cm} (A.2.4)

Hence, \(\bar{e}\) is defined so that the \(e_i\)'s sum to zero. To construct the \(m^{th}\) bootstrap data set, let \(\{e_i^{(m)}\}_{i=1}^{n}\) be sampled with replacement from the residuals and then let

\[Y^{(m)}_i = \mu(Z_i, \hat{\beta}) + \hat{\sigma} \gamma(Z_i, \hat{\beta}, \hat{\theta}) e_i^{(m)}\]  
\hspace{1cm} (A.2.5)

here \(e_i^{(m)}\) is not the residual from the \(i^{th}\) of the original observations but is equally likely to be any of the \(n\)-residuals from the original observations.
A.2.2 Bootstrapping Measurement Error Model

In a measurement error problem, a typical data vector consists of $Z_i$ the error free covariates and a subset of the following data: the response $Y_i$, the true covariates $X_i$, 
\[ \{w_{i,j} : j = 1, \ldots, k_i \} \] which are replicate surrogates for $X_i$. Here the total collection of data is divided into homogeneous data sets, which have the same variables measured on each observation and are from a common source, for example, primary data, internal data, external replication data, and internal validation data. The method of resampling pairs ignores the various data subsets. Here resampling is done by with replacement and independently from each of the homogeneous data sets. This ensures that each of the bootstrap data set has the same amount of validation data as the actual data set.

Resampling residuals is applicable to validation data when there are two regression models: one for $Y_i$ given $(X_i, Z_i)$ and another for $w_i$ given $(X_i, Z_i)$. One fits both models and resamples residuals from the first to create the bootstrap $Y_i^{(m)}$'s and from the second to create the $w_i^{(m)}$'s. This method generates sampling distributions that are conditional on the observed $(X_i, Z_i)$'s. The parametric bootstrap can be used when the response, given the observed covariates, has a distribution in a known parametric family. Suppose one has a logistic regression model with internal validation data. One can fix the $(X_i, Z_i, w_i)$ vectors of the validation data and can create bootstrap responses using $(X_i, Z_i)$ in place of $Z_i$. Since $w_i$ is a surrogate and cannot be used to create the bootstrap responses of validation data. For the non-validation data, one fixes the $(Z_i, w_i)$ vectors and approximates the model for $Y_i$ given $(Z_i, w_i)$ and then creates bootstrap responses distributed according to the fitted model.
A.2.3 Bootstrap Confidence Intervals

Standard parametric confidence intervals can provide a measure of significance for regression coefficients. Yet they require acceptance of Gaussian assumptions regarding estimates of coefficients for their validity. Diagnostic analysis did not support these assumptions, especially given the limited data available to estimate the variability from the multitude of sources. Alternatives to the standard parametric confidence intervals are the semiparametric or nonparametric methods using bootstrap estimates of the variability of the coefficient estimates. If the analysis requires nonparametric bootstrap percentile confidence intervals to infer the observed significance level of the effects; the multiple linear regression is performed with 1000 bootstrap replications, by fixing the design matrix and resampling from the possible responses conditional on each treatment combination. The bootstrap distribution of each regression coefficient is compiled, and the 5th and 95th percentiles of the empirical distribution formed the limits for the 95% bootstrap percentile confidence interval. If the confidence interval failed to include 0, then the p-value is deemed to be less than or equal to 0.05, and the effect is said to be significant.

A.3 Sandwich Method

If anyone considers inferences about a parameter vector $\Theta$, they assume that the estimate $\hat{\Theta}$ maximizes an estimating criterion, $l(\Theta)$, which is the working log likelihood of an actual density function. Define $\exp(l) = \exp\left(\sum_{i=1}^{n} l_i\right)$ as the quasi likelihood function, hence $l_i$ is the log quasi likelihood for the $i^{th}$ case and $l$ is the log quasi likelihood for the entire data set. The estimating equation, score function becomes,

$$\Delta_i(\Theta) = \frac{\delta}{\delta \Theta} l(\Theta | \tilde{y}_i)$$
The score covariance matrix is obtained as,

\[ \Gamma_n = \sum_{i=1}^{n} E \left\{ \Delta_i(\Theta) \Delta_i(\Theta)' \right\} \]  \hspace{1cm} (A.3.1)

The negative expected Hessian is defined as,

\[ H_n = -\sum_{i=1}^{n} E \left[ \frac{\delta}{\delta \Theta'} \Delta_i(\Theta) \right] \]  \hspace{1cm} (A.3.2)

If \( l \) is the true log likelihood, then \( H_n = \Gamma_n \), but it fails for quasi likelihood since the herein the parameter \( \Theta \) is determined by the equation \( E \left\{ \Delta_i(\Theta) \right\} = 0 \) for all \( i \) or possibly through the weaker constraints, \( \sum E \left\{ \Delta_i(\Theta) \right\} = 0 \).

Partitioning \( \Theta \) as; \( \Theta = \left( \Theta^{(1)} \Theta^{(2)} \right)' \), where the dimension of \( \Theta^{(1)} \) is \( p \) and it is the vector of interest and the other one \( \Theta^{(2)} \) is the vector of nuisance parameters. Similarly partitioning \( H_n \), it becomes,

\[
H = \begin{pmatrix}
H_{11} & H_{12} \\
H_{21} & H_{22}
\end{pmatrix}
\]

Also define, \( H_{11,2} = H_{11} - H_{12} H_{22}^{-1} H_{21} \).

Let \( \hat{\Theta}_0 = \left( \Theta^{(1)}_0 \Theta^{(2)}_0 \right)' \) denote the maximum likelihood estimate subject to \( \Theta^{(1)} = \Theta^{(1)}_0 \); under large sample distribution of the log quasi likelihood ratio becomes,

\[
\ell \left( \Theta^{(1)}_0 \right) = 2 \left\{ l(\hat{\Theta}) - l(\hat{\Theta}_0) \right\}
\]
**Theorem**

If \( \Theta_{(i)} = \Theta_{0(i)} \), then as the number of independent observations increases, \( \ell\left( \Theta_{0(i)} \right) \) converges in distribution to \( \sum_{k=1}^{p} \lambda_k \kappa_k \), where \( \kappa_1, \ldots, \kappa_p \) are independently distributed as \( \chi^2 \), and \( \lambda_1, \ldots, \lambda_p \) are the eigen values of \( H_{11,2} = H_{11} - H_{12} H_{22}^{-1} H_{21} \).

Using this result to perform the quasi likelihood ratio test, \( H_0 : \Theta_{(i)} = \Theta_{0(i)} \) or to compute a quasi likelihood confidence set for \( \Theta_{0(i)} \) estimate the matrices \( H \) and \( \Gamma \). If the data are independent, the approach is to replace the theoretical expectation in (A.3.1) and (A.3.2) with the empirical averages. Since chi-squared random variables are easy to generate, use the simulated data from the distribution of \( \sum_{k=1}^{p} \lambda_k \kappa_k \) to find the quantiles of the distribution of \( \sum_{k=1}^{p} \lambda_k \kappa_k \).

**A.3.1 An Example: Sandwich Formula and Linear regression**

Consider an ordinary multiple regression without measurement error, \( Y_i = \beta_0 + \beta'_z Z_i + \epsilon_i \), where the \( \epsilon \)’s are independent, mean-zero random variables. Let \( Z_i^* = \begin{pmatrix} 1 & Z_i' \end{pmatrix} \) and

\( \Theta = \begin{pmatrix} \beta_0 & \beta'_z \end{pmatrix}' \). Then the ordinary least square estimator is an M-estimator with

\( \varphi_i (Y_i, \Theta) = \left( Y_i - \beta_0 - \beta'_z Z_i \right) Z_i^* \). Also,

\[
\frac{\delta}{\delta \Theta'} \varphi_i (Y_i, \Theta) = -Z_i^* \left( Z_i^* \right)' \tag{A.3.3}
\]

And

\[ A_n = -\frac{1}{n} \sum_{i=1}^{n} Z_i^* \left( Z_i^* \right)' \tag{A.3.4} \]

Assume the variance of \( \epsilon \) is a constant \( \sigma^2 \) for all \( i \), then
\[ B_n = -\sigma^2 A_n \] (A.3.5)

Generally, \( \sigma^2 \) is not known and is estimated by the residual mean square. The sandwich formula uses, \( \hat{A}_n = A_n \) and

\[ \hat{B}_n = \frac{1}{n} \sum_{i=1}^{n} \left( Y_i - \hat{\beta}_0 - \hat{\beta}_i Z_i \right) Z_i^* \left( Z_i^* \right)' \] (A.3.6)

Here no assumption is made about the distribution of \( \varepsilon_i \) but assume homoscedasticity; now consider the heteroscedastic model with \( \text{Var}(\varepsilon_i) = \sigma_i^2 \) depends on \( Z_i \). Then (A.3.5) becomes,

\[ B_n = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 Z_i^* \left( Z_i^* \right)' \] (A.3.7)

The estimated value of (A.3.7) can be obtained from (A.3.6). Therefore, the sandwich formula is heteroscedasticity consistent. The inefficiency of the sandwich estimator can be seen in this example.

A.4 Generalized Linear Models

Exponential families have density or mass function

\[ f(y | x, z, \Theta) = \exp \left\{ \frac{y \eta - A(\eta)}{\Phi} + c(y, \Phi) \right\} \] (A.4.1)

The mean and variance of \( Y \) are given as \( A'(\eta) \) and \( \Phi A''(\eta) \) respectively. If \( \eta \) is a function of linear combination of predictors, say \( \eta = X_i (\xi) \) where \( \xi = \beta_0 + \beta_1 X + \beta_2 Z \); then we have a generalized linear model. Generalized models include many of the common regression models, for instance, normal, logistic, Poisson, and gamma, etc. generalized models are mean and variance models in the observed data, and can be fit using QVF methods.
If define \( L = (A' \circ X_i)^{-1} \), then \( L(\mu) = \xi \); hence \( L \) is called the link function since it links the mean of the response and the linear predictor, \( \xi \). If \( X_i \) is the identity function, then the model is canonical; that implies \( L = (A')^{-1} \) which is called canonical link function. The link function \( L \), or equivalently \( X_i \), should be chosen so that the model fits the data as well as possible. If the canonical link function fits reasonably well, then it is typically used, because it simplifies the analysis.

A.5 The Idea of Wilks’s Theorem

To realize the importance of Wilks’s theorem consider the parametric likelihood. Let \( L(\theta) \) denote the classical likelihood of a sample, \( \hat{\theta} \) is the typical maximum likelihood estimator and the likelihood ratio becomes,

\[
R(\theta) = -2 \log \left( \frac{L(\theta)}{L(\hat{\theta})} \right)
\]

(A.5.1)

Let \( \theta_0 \) be the true value of \( \theta \) and assume that there is no nuisance parameters and suppose \( q \) be the rank of the asymptotic variance matrix of \( \frac{\hat{\theta}}{n} \). Wilks’s theorem states that under appropriate regularity conditions, \( R(\theta_0) \) has an asymptotic \( \chi^2_q \)-distribution. This result is the key to constructing parametric likelihood based confidence regions which is defined as,

\[
\Pr \left[ \chi^2_q \leq r_0 \right] = 1 - \alpha
\]

(A.5.2)

Here \( r_0 \) is some threshold value and \( (1 - \alpha) \) is the desired nominal coverage of the region. Then the appropriate region is,
\[ \mathcal{Z}_{r_0} = \{ \theta : R(\theta) \leq r_0 \} \quad (A.5.3) \]

In view of Wilks’s theorem, the asymptotic coverage of \( \mathcal{Z}_{r_0} \) equals \( (1 - \alpha) \) as \( n \to \infty \):

\[ \Pr[\theta_0 \in \mathcal{Z}_{r_0}] = \Pr(R(\theta_0) \leq r_0) \to (1 - \alpha) \quad (A.5.4) \]

A major property of empirical likelihood is that it admits a nonparametric version of Wilks’s theorem. Therefore the above procedure can be employed in empirical likelihood inference. More specifically, define the log likelihood ratio \( R(\theta_0) \), and threshold value \( r_0 \) as at \( (A.5.1) \) and \( (A.5.2) \) respectively; construct the confidence region according to \( (A.5.3) \). Then \( (A.5.4) \) holds by the empirical likelihood version of wilks’s theorem. Now in empirical likelihood context \( \theta \) is a function of the population mean, say \( \theta = h(\mu) \). In this case the estimator, \( \hat{\theta} = h(\bar{X}) \) and it the function of the sample mean \( \bar{X} = \frac{1}{n} \sum X_i \). Here the rank of \( \theta \) is \( q \), and \( X_i \) and \( \mu = (\mu^{(i)}, \ldots, \mu^{(s)})' \) are of length \( s \). Let \( \mu_0 = E(X) \) is the true value of the population mean, and let \( \theta_0 = h(\mu_0) \) be the true value of \( \theta \). If \( h = (h^{(1)}, \ldots, h^{(q)})' \) has a continuous derivative in a neighborhood of \( \mu_0 \) then the asymptotic variance matrix of \( \frac{\hat{\theta}}{n} \) is

\[ V = \nu_0 \Sigma \nu_0' \], where \( \nu_0 = (\nu_0^{(ij)}) \) denotes \( q \times s \) matrix defined by,

\[ \nu_0^{(ij)} = \frac{\partial h^{(i)}(\mu)}{\partial \mu^{(j)}} |_{\mu = \mu_0} \]

and \( \Sigma = E \left[ (X - \mu_0)(X - \mu_0)' \right] \) is the population variance matrix.
Wilks’s Theorem

Assume $X$ has finite variance and $h(\cdot)$ has a continuous derivative in a neighborhood of $\mu_0$. Let $t \leq \min(r, s)$ denote the rank of $V$ and let $R(\cdot)$ be the empirical likelihood function. Then $R(\theta_0)$ has an asymptotic $\chi^2_t$ distribution.

A.6 Model Identifiability

In some statistical models, different parameter values can give rise to identical probability distributions. When this happens, there will be a number of different parameter values associated with the maximum likelihood of any set of observed data. This is referred to as the model identifiability problem. For example, suppose someone attempts to compute the regression equation predicting $Y$ from three variables $X_1, X_2,$ and $X_3$ where $X_3 = (X_1 + X_2)$ the program will probably crash or give an error message because it cannot find a unique solution.

The model is the same if $Y = X_1 + 2X_2 + 3(X_1 + X_2)$, or $Y = X_1 + 1.5X_2 + (X_1 + X_2)$, etc. indeed there are an infinite number of equally good possible solutions. A model is identifiable if the parameter values uniquely determine the probability distribution of the data and the probability distribution of the data uniquely determines the parameter values. Let $\Theta$ be the parameter value of the model and $Y$ be the observed data and $F(y; \Theta)$ be the probability distribution of the data.

A model is identifiable if for all $(\Theta_0, \Theta) \in \Omega$ and for all $y \in \mathbb{N}_Y$:

$$F(y; \Theta_0) = F(y; \Theta) \text{ if and only if } \Theta_0 = \Theta \quad \text{(A.6.1)}$$

where $\Omega$ denotes the set of all possible parameter values, and $\mathbb{N}_Y$ is the set of all possible values of the data. The most common cause of model nonidentifiability is a poorly specified model. If the number of unique model parameters exceeds the number of independent pieces of observed information, the model is not identifiable. If the model is not identifiable, one can make
it so by imposing various constraints upon the parameters. When there appears to be sufficient
total observed information for the number of estimated parameters, it is also necessary to specify
the model unambiguously. It is difficult to specify general conditions that are sufficient to
guarantee (global) identifiability. Fortunately, it is fairly easy to determine local identifiability.
One can require that the columns of the Jacobian matrix, the first order partial derivative of the
likelihood function with respect to the unique model parameters, are independent. Alternatively,
one can examine whether the Fisher information matrix possesses eigenvalues greater than zero;
these two approaches are equivalent. A standard practice for checking local identifiability
involves using multiple sets of initial values for parameter estimation. Different sets of initial
values that yield the same likelihood maximum should result in the same final parameter
estimates. If not, the model is not locally identifiable. When applying a nonidentifiable model,
different people may draw different conclusions from the same model of the observed data.
Before one can meaningfully discuss the estimation of a model, model identifiability must be
verified. If researchers come up against identifiability problems, they can first identify the
parameters involved in the lack of identifiability from their extremely large asymptotic standard
errors, and then impose reasonable constraints on identified parameters based on prior
knowledge or empirical information.

A.7 Appendix to Chapter 3 and 4

A.7.1 naive <- glm(cbr ~ dust + smoking + expo,
        family= binomial,
        data =simex_data,
        x=TRUE, y=TRUE)

A.7.2 mod.sim<-simex(naive, SIMEXvariable="dust",
        measurement.error=2,
        fitting.method = "quadratic")
A.7.3

elm<-function( x, mu, lam, maxit=25, gradtol=1e-7, 
        svdtol = 1e-9, itertrace=F ){
    x = as.matrix(x)
    n = nrow(x)
    p = ncol(x)
    mu= as.vector(mu)
    if( length(mu) != p )
        stop("Mu must have same dimension as observation vectors.")
    if( n <= p )
        stop("Need more observations than variables in elm.")
    z = t( t(x) -mu )
    TINY = sqrt( .Machine$double.xmin )
    scale = mean( abs(z) ) + TINY
    z = z/scale
    if( !missing(lam) ){
        lam=as.vector(lam)
        lam=lam*scale
        if( logelr(z,rep(0,p),lam)>0 )lam=rep(0,p)
    }
    if( missing(lam) )
        lam=rep(0,p)
    if( svdtol < TINY )svdtol = TINY
    if( gradtol < TINY)gradtol = TINY
    nwtswt0 = c( 3^-c(0:3), rep(0,12) )
    gwts = 2^(-c(0:(length(nwtswt0)-1)))
    gwts = (gwts^2 - nwtswt0^2)^.5
    nits = 0
    gsize = gradtol + 1
    while( nits<maxit && gsize > gradtol ){
        arg  = 1 + z %*% lam
        wts1 = as.vector( llogp(arg, 1/n) )
        wts2 = as.vector( -llogpp(arg, 1/n )^.5
        grad = as.matrix( -z*wts1 )
        grad = as.vector( apply( grad, 2, sum ) )
        gsize= mean( abs(grad) )
        hess = z*wts2
        svdh = svd( hess )
        if( min(svdbh$d) < max(svdbh$d)*svdtol )
            svdh$d = svdh$d + max(svdbh$d)*svdtol
        nstep = svdh$d %*% (t(svdbh$v)/svdh$d)
        nstep = as.vector( nstep %*% matrix(wts1/wts2,n,1) )
        gstep = -grad
        if( sum(nstep^2) < sum(gstep^2) )
gstep = gstep*sum(nstep^2)^.5/sum(gstep^2)^.5
ologelr = -sum( llog(arg,1/n) )
ninner = 0
for( i in 1:length(nwts) ){
    nlogelr = logelr( z,rep(0,p),lam+nwts[i]*nstep+gwts[i]*gstep )
    if( nlogelr < ologelr ){
        lam = lam+nwts[i]*nstep+gwts[i]*gstep
        ninner = i
        break
    }
}
nits = nits+1
if( ninner==0 ) nits = maxit
if( itertrace )
    print( c(lam, nlogelr, gsize, ninner) )

list( logelr=nlogelr, lambda = lam/scale, grad=grad*scale,
    hess=t(hess)%*%hess*scale^2, wts=wts1, nits=nits )
}

logelr=function( x, mu, lam ){
x = as.matrix(x)
n = nrow(x)
p = ncol(x)
if( n <= p )
    stop("Need more observations than variables in logelr.")
mu= as.vector(mu)
if( length(mu) != p )
    stop("Length of mean doesn't match number of variables in logelr.")
z = t( t(x) -mu )
arg = 1 + z %*% lam
- sum( llog(arg,1/n) )
}
llog=function( z, eps ){
    ans = z
    lo = (z<eps)
    ans[ lo ] = log(eps) - 1.5 + 2*z[lo]/eps - 0.5*(z[lo]/eps)^2
    ans[ !lo ] = log( z[!lo] )
    ans
}
llogp=function( z, eps ){
    ans = z
    lo = (z<eps)
    ans[ lo ] = 2.0/eps - z[lo]/eps^2
    ans[ !lo ] = 1/z[!lo]
    ans
\[
\text{logpp=function( z, eps )}{ \\
\text{ans = z} \\
\text{lo = (z<eps)} \\
\text{ans[ lo ] = -1.0/eps^2} \\
\text{ans[ !lo ] = -1.0/z[!lo]^2} \\
\text{ans} \\
}\]

A.7.4a Upper Bound Function

\[
n <- \text{nrow(X)} \\
p <- \text{ncol(X)} \\
df<-p \\
\text{tol <- matrix(rep(1e-05,df),nrow=p)} \\
\text{cut <- qchisq(.95,p)} \\
x.bar <- \text{matrix(c(mean(X[,1]),mean(X[,2])), nrow=p)} \\
t2 <- \text{matrix(c(max(X[,1]), max(X[,2])),nrow=p)} \\
ds <- \text{matrix(1/n,nrow=n)} \\
t1 <- x.bar \\
dif <- t2-t1 \\
\]

while(any(dif>tol))
{
\text{tau=(t1+t2)/2} \\
\text{ELM=elm(X, beta0, tau, maxit=25, gradtol=1e-7, svdtol = 1e-9, itertrace=F )} \\
\text{elratio=ELM$logelr} \\
\text{if(elratio>cut) t2=tau} \\
\text{if(elratio<=cut) t1=tau} \\
\text{dif=t2-t1} \\
}\]

\text{UB<- (t1+t2)/2}

A.7.4.1b Lower bound function

\text{t3<- matrix(c(min(X[,1]), min(X[,2])),nrow=p)} \\
\text{dif.L<-t1-t3} \\
\]

while(any(dif.L>tol)){
\text{tau.L=(t1+t3)/2} \\
\text{ELM.L=elm(X, beta0, tau.L, maxit=25, gradtol=1e-7, svdtol = 1e-9, itertrace=F )} \\
\text{elratio.L=ELM.L$logelr} \\
\text{if(elratio.L>cut) t3=tau.L} \\
\text{if(elratio.L<=cut) t1=tau.L} \\
\text{dif.L=t1-t3} \\
}\]
\text{LB<- (t1+t3)/2}