An Abstract of the Thesis of

Robert Delongchamp for the degree of Doctor of Philosophy in Statistics presented on February 18, 1993.

Title: Analysis of Epidemiological Data with Covariate Errors

Abstract approved: Redacted for Privacy

Donald A. Pierce

In regression analysis, random errors in an explanatory variable cause the usual estimates of its regression coefficient to be biased. Although this problem has been studied for many years, routine methods have not emerged. This thesis investigates some aspects of this problem in the setting of analysis of epidemiological data.

A major premise is that methods to cope with this problem must account for the shape of the frequency distribution of the true covariable, e.g., exposure. This is not widely recognized, and many existing methods focus only on the variability of the true covariable, rather than on the shape of its distribution. Confusion about this issue is exacerbated by the existence of two classical models, one in which the covariable is a sample from a distribution and the other in which it is a collection of fixed values. A unified approach is taken here, in which for the latter of these models more attention than usual is given to the frequency distribution of the fixed values.

In epidemiology the distribution of exposures is often very skewed, making these issues particularly important. In addition, the data sets can be very large, and another premise is that differences in the performance of methods are much greater when the samples are very large.

Traditionally, methods have largely been evaluated by their ability to remove bias from the regression estimates. A third premise is that in large samples there may be various methods that will adequately remove the bias, but they may differ widely in how nearly they approximate the estimates that would be obtained using the unobserved true values.

A collection of old and new methods is considered, representing a variety of basic rationales and approaches. Some comparisons among them are made on

theoretical grounds provided by the unified model. Simulation results are given which tend to confirm the major premises of this thesis. In particular, it is shown that the performance of one of the most standard approaches, the "correction for attenuation" method, is poor relative to other methods when the sample size is large and the distribution of covariables is skewed.

Analysis of Epidemiological Data with Covariate Errors

by

Robert Delongchamp

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed February 18, 1993

Commencement June 1993

APPROVED:

Redacted for Privacy

Professor of Statistics in charge of major

Redacted for Privacy

Head of department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented February 18, 1993

Prepared by Robert Delongchamp

Acknowledgments

Over years, many friends and acquaintances have shaped this work; far too many to attempt an enumeration. Some were teachers, others gave insight and many gave encouragement to stay the course. All of these contributions are valued and appreciated.

Don Pierce through his guidance and our discussions has greatly influenced this thesis and, more importantly, my overall perception and understanding of statistics. Equally appreciated are his efforts to bolster my morale at those times when that was most needed.

The faculty, students and staff of the Department of Statistics made the time spent in Oregon educationally fruitful as well as very enjoyable. I thank Qing Liu and Gen Downing for their assistance with the final logistics of completing this thesis.

The love, patience and understanding of my family, Gerry Ann, Tracey, Troy and Travis, was especially helpful. My parents, Roy and Shirley, deserve special recognition for nurturing my curiosity.

The financial support provided by Dow Corning Corporation is gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

Analysis of Epidemiological Data with Covariate Errors

Chapter 1. Introduction

There has been some attention to error-in-variables problems over many years, but the general problem is difficult and little of what has been developed has come into routine use. In recent years, the general problem has received more attention due to requirements in epidemiological and biostatistical applications. The statistical models involved there differ from those treated in the bulk of the error-in-variables literature, and attempts to accommodate measurement error into these models has caused a reconsideration of the issues. Although research on measurement error models has had a long history with a voluminous literature (Fuller, 1987), the majority of these writings pertain to models that assume Gaussian settings for both the regression and the errors-in-variables. These models usually do not satisfy the needs in biostatistics, especially in epidemiology. Perhaps somewhat ironically, consideration of more general settings, even though they are more difficult, seems to have made some of the basic issues clearer.

Traditionally, models for covariate error have been categorized as structural or functional. This dichotomy pertains to whether the true covariate values are assumed to come from a distribution, structural model, or whether the true covariate values are assumed to be fixed, functional model. The distinction of applicability of these two models is unclear. Part of point here is that there may be less distinction than thought, and structural type reasoning is important even in the functional setting. That is, the unobserved empirical distribution of the true covariate values in the functional model plays an analogous role to the distribution of true covariate values in a structural model. Pierce *et al.* (1990, 1992) introduced the quasi-structural model which is just the structural model based upon the empirical distribution of the true covariate values. In the classical functional model, the set of true covariate values are essentially incidental parameters. In the quasi-structural approach, the empirical distribution of the true covariate values is taken as the parameter. This subtle distinction is quite important, especially in large samples, because information regarding the empirical distribution accumulates nicely, whereas, the information about incidental parameters does not.

The primary thesis underlying this dissertation involves the following rationale and conjectures:

1. It is important that methods explicitly and carefully consider the distribution of the true covariate values. In the functional setting, the empirical distribution can play this role through the quasi-structural model.
2. This point is far more important in large samples than in small ones, and hence critical in epidemiological studies, which typically involve very large samples.
3. It is inadequate to focus only on how well an estimator removes the bias present in the naive estimates. There will be various methods that adequately remove this bias, but they may differ greatly in their "efficiency"; that is, how closely they come to the estimates based upon knowing the true covariate values.
4. In particular as the sample size increases, it seems plausible that the classical "correction for attenuation" approach will be increasingly less efficient relative to the best possible solution.

This dissertation lays out some methods that represent somewhat different approaches to measurement error and considers how these approaches seem to fit in with the basic thesis. Conclusions stemming from these considerations are confirmed through some limited simulation. The simulations are not intended to be comprehensive because there are simply too many factors to consider. Rather the aim was to select a few scenarios that clarify whether the underlying rationales of the thesis hold up.

In this work, the general form for a measurement error model has three component distributions. Because it is central to the thesis, the distribution of the true covariate values has already been discussed. The other two distributions are the conditional distribution of the response given the true covariate and the conditional distribution of the observed covariate given the true covariate. In epidemiological applications, measurement error models mainly need to consider a generalized linear model for the response variable, and frequently a multiplicative form for the covariate errors. However, elaborate consideration of the differences between such measurement error models and Gaussian regression models with additive covariate errors is less fundamental than consideration of the highly non Gaussian distributions

of true covariate values and the very large sample sizes that often characterize epidemiological data. For example, the distinction between classical models and those involving a generalized linear model with multiplicative errors can often be adequately dealt with simply by minor generalizations based on second moment structures.

Concerning the conditional distribution of the observed covariate, the approach here is to assume that it is tentatively specified. In practice it will ordinarily not be fully known, but the complexity of the general measurement error problem without this specification distracts attention from other important aspects studied here. There will be problems with model identifiability in any approach that attempts a treatment with no restrictions on both the distribution of the true covariates and the conditional distribution of the measurement errors. As a practical recourse, this identifiability problem is often dealt with through specific assumptions about the conditional distribution of the observed covariate. The orientation of this research has been to explore the modeling issues that remain in this situation, i.e., the assumption that the conditional distribution of the unobserved covariates is completely specified.

The formulation and especially the simulation has been influenced by needs in the analysis of cancer incidence among the A-bomb survivors, but many of the features are common to epidemiological studies. Mainly, these are highly skewed distributions of true covariates (exposures), very large sample sizes, and responses which are individually rather "imprecise" in the sense that even if the regression parameters were known, each outcome would provide little information about the corresponding covariate error. In this context, methods that replace the observed covariates in the naive regression with estimates of the expected value of the true covariate given the observed covariate can be very efficient as well as being easy to implement. When the distribution of the true covariates is not Gaussian, such a model will be quite different from most of the classical approaches.

Chapter 2 states the modeling assumptions and discusses some of the traditional techniques for handling these assumptions. Under a structural setting for the distribution of the true covariates, the maximum likelihood estimator and the iteratively reweighted least squares estimator are explained. These standard estimation methods are related to each other and to replacement methods in general as well as two recently introduced replacement methods by Whittemore (1989) and Schafer (1990). In a functional setting, the generalization of classical correction for

attenuation methods to the generalized linear regression setting is explored and a simple extension presented. The quasi-structural model, the structural model based upon the empirical distribution of the true covariate values, is introduced and it is shown that the relative efficiency of the traditional approaches can be interpreted in the context of this model. A method by Nakamura (1990) is discussed. This method is pursued because it can be viewed both as an extension of "correction for attenuation" concepts and as an approximation to the quasi-structural maximum likelihood estimator.

Chapter 3 presents some simulation results intended to explore the four points itemized above. The usual practice of simply studying the removal of bias from the naive estimates is not adequate for large samples and an "efficiency" criterion is presented which monitors how close an estimator is to the estimate that would be obtained without measurement errors. The efficiencies of several estimators relative to quasi-structural maximum likelihood estimator are demonstrated in a simulation experiment. The approximation to the quasi-structural maximum likelihood score implied by a method's estimating equation can qualitatively predict its relative efficiency. These predictions incorporate many of the important characteristics of cohort studies.

In Chapter 4, the statistical methods in Chapter 2 are presented in some detail, including precise documentation of the simulation methods. Somewhat unconventionally the conclusions are given at the end of Chapter 3, since Chapter 4 deals with several simulation details that do not seem suitably placed in an appendix.

This work is an extension of work by Pierce *et al.* (1990, 1992) where the quasi-structural formulation was first introduced and applied to the A-bomb survivor data. It has been heavily influenced by the methods presented by Amemiya and Fuller (1988), Nakamura (1990), Schafer (1990) and Whittemore (1989). In addition, the notation and modeling perspective has been influenced by the discussion of empirical Bayes methods in the book by Maritz and Lwin (1989).

Chapter 2. Measurement Error Models and Estimation Methods

The primary aim of this chapter is to lay out a modeling strategy and to establish notation. Several modeling approaches and estimation methods are presented and they provide the foundation for subsequent discussions of some issues that arise whenever there are measurement errors in a regression covariate.

The basic problem is to estimate the regression coefficient in the relationship between the expected value of a response, y_i, and the value of an explanatory variable, x_i, that is indirectly observed through another covariate, z_i. The observed covariate, z_i, is the true x_i contaminated by a random error.

There are two customary approaches to deal with the unobserved x_i. One approach is to assume that x_i is a random observation from a probability distribution, $F_X(\cdot)$. Such models are referred to as structural models for measurement error. The other approach is to assume that the true value of x_i is fixed but unknown, thus essentially a parameter, and such models are referred to as functional models for measurement error.

Consideration will also be given to another approach to the functional setting that is based on the quasi-structural model (Pierce *et al.*, 1992). In its construction this model is the structural model that results when the empirical distribution of the true vector, $\mathbf{x}$, replaces $F_X(\cdot)$. As will be shown, the quasi-structural model is a useful construct from which to develop and evaluate estimation methods. The relevance of careful attention to the empirical distribution of $\mathbf{x}$, even though it is not directly observed, will be emphasized in this work.

The role that is played by the empirical distribution can be seen by evaluating the expected value of x given z for grouped data. Pierce *et al.* (1990) presented the cross tabulation of hypothetical data that has been duplicated in Table 1. In this table the data satisfy $E_{Z|X=x}(z) = x$. It makes sense to think of $E_{X|Z}(x)$ even though there may otherwise be little reason to think of x as a random variable; grouping helps illustrate this. The main point is that this calculation, i.e., Avg(x|z), has value and meaning in a functional setting. It is noteworthy that the influence of the marginal distribution of the true doses has little to do with grouping. Grouping is simply a device which facilitates the recognition that what is important is the relative frequency of the true values, i.e., the empirical distribution. The quasi-structural

model formalizes this aspect of the computation of similar conditional expectations in the measurement error problem.

Several established methods for estimating the regression parameters in measurement error models are introduced and their connection to the quasi-structural model are examined. For a structural model and a quasi-structural model, the maximum likelihood estimates and the iteratively reweighted least squares estimates, which may be thought of as approximations to these, are discussed. Under a functional setting two methods are considered that attempt to construct consistent estimates of the regression parameters; the first method expands on the correction for attenuation method (Fuller, 1987), and the second method corrects the naive score equation (Nakamura, 1990). Two other methods, by Whittemore (1989) and Schafer (1990), are examples of replacement methods. That is, they assume a distribution for x to estimate the expected value of the unobserved x_i given an observed value, z_i. Then this estimate supplants the unobserved x_i in a conventional regression analysis. Strictly speaking, replacement methods represent a structural approach but they can also be thought of as an empirical Bayes solution to the estimation problems originating with the incidental parameters, $\mathbf{x}$, in a functional model.

Table 1. Artificial Example Indicating Basic Concepts

True dose	Estimated dose z (Gy)						Number of
x (Gy)	1	2	3	4	5	6	Survivors
-	-						-
-	-	-					-
2	250	500	250				1000
3		75	150	75			300
4			33	66	33		132
5				15	30	15	60
-					-	-	-
Avg(x|z)	-	-	2.50	3.62	-	-	

2.1 Class of Models

The measurement error models that will be studied are determined by a class of probability distributions for y_i given x_i in combination with a class of probability distributions for z_i given x_i. Since, conditional on x_i, y_i and z_i are assumed to be independent, these two classes can be defined separately. This research is primarily concerned with the functional setting; that is a setting in which the unknown true values of the covariate are fixed. However, the importance of considering the empirical distribution of the values, $\mathbf{x}$, is a dominant theme and many of the models to be discussed are structural models, i.e., a distribution for x is assumed.

2.1.1 Distribution of *x*

The distribution of the unobserved covariates will be unknown and most of the methodology that will be discussed is oriented toward estimating the regression under this assumption. In measurement error models, three distributions for x are distinguished by notation in the sequel. In its definition, a structural model for measurement errors specifies that the values of the unobserved covariate are a sample from a distribution. That is, $x_i \sim \mathrm{F}_X(\cdot)$, $i = 1, \cdots n$. If $\mathrm{F}_X(\cdot)$ is assumed to be known in a structural model, standard statistical methods such as maximum likelihood estimation or weighted least squares estimation are, at least in concept, straightforward to carry out. In a functional setting, the empirical distribution, $\mathrm{F}_X^n(\cdot)$, can play an analogous role to $\mathrm{F}_X(\cdot)$ in a structural setting.

Because the true distribution, either $\mathrm{F}_X^n(\cdot)$ or $\mathrm{F}_X(\cdot)$ will be unknown, notation is adopted to emphasize the dependence of estimates on an assumed distribution, $\mathrm{G}(\cdot)$. This distinction between the true distribution and an assumed distribution, $\mathrm{G}(\cdot)$, is conveyed through notation, e.g., $\mathrm{f}_{Y|UZ}(\cdot;\mathrm{G})$, which is intended to indicate that $\mathrm{F}_X(\cdot)$ is a parameter that has been estimated by $\mathrm{G}(\cdot)$. To facilitate the later use of this notation, $\mathrm{F}_X(\cdot)$ will be included in the argument list of functions that depend upon its specification. The practicality of this notation should be obvious in situations where several estimators, that differ only in the adopted $\mathrm{G}(\cdot)$, are being compared. For example, the estimator of φ that maximizes the likelihood based upon a given $\mathrm{G}(\cdot)$ will be denoted as $\hat{\varphi}(\mathrm{G})$. Because $\mathrm{F}_X(\cdot)$ is simply replaced by $\mathrm{G}(\cdot)$, the resultant likelihood can be considered to be a pseudo-likelihood (Gong and Samaniego, 1981).

2.1.2 Distribution of z

The measurement errors are characterized by the conditional distribution, $f_{Z|X}(\cdot)$. The approach taken here is to tentatively assume that this density and the values of any parameters are known. In practice this assumption is very unlikely to be met. The foremost interest in this thesis is on issues that surround the modeling of the distribution of $\mathbf{x}$ and without this assumption or something similar the theory becomes cluttered because the models tend to be unidentifiable. This clutter detracts from the important issues to be discussed and, to avoid this aspect of the measurement error problem, it is convenient to take $f_{Z|X}(\cdot)$ as given.

As a practical recourse to identifiability problems, the assumption that $f_{Z|X}(\cdot)$ and its parameters are specified is common, although not universal, in measurement error models. It is a standard assumption in some of the classical Gaussian models where the mean of the measurement error distribution is taken to be x and the variance known (Cochran, 1968; Kendall and Stuart, 1979; Fuller, 1987). Many recently published methods, which have been proposed for epidemiological data, such as the models discussed by Armstrong (1985), Burr (1988), Carroll *et al.* (1984), Nakamura(1990, 1992), Prentice (1982), Stefanski and Carroll (1985), Schafer (1990) and Whittemore (1989), make similar, if not the same, distributional assumptions. In real data situations, a sensitivity analysis can give an assessment of the practical significance of this assumption (Pierce *et al.*, 1992).

In modeling measurement errors, it is natural to assume that $E_{Z|X=x}(z) = x$. Although technically unnecessary, this assumption simplifies the form of many of the estimators that are discussed and $f_{Z|X}(\cdot)$ can usually be defined so that $E_{Z|X=x}(z) = x$. The simplification occurs because the assumption implies the relationship, $E_Z(z) = E_X(x)$, making $\bar{z}$ an unbiased estimator of $E_X(x)$.

In modeling the variance, the notion of additive or multiplicative error differentiates two classes of measurement error distributions. In particular, assume that there is a distribution, $f_{\mathbf{e}}(\cdot)$, with $\mathbf{e}$ being a sample of independent observations from that distribution. Then the additive error model has $\mathbf{z} = \mathbf{x} + \mathbf{e}$, and the multiplicative error model has $\mathbf{z} = \mathbf{x} \times \mathbf{e}$. Thus the terms, additive and multiplicative, denote a scale on which the measurement errors have constant variance, i.e., the

variance is independent of x. Additive models have received the most consideration in related work. The emphasis in this thesis will be on multiplicative errors.

2.1.3 Distributions for *y*

The distribution of y defines an underlying regression relationship. The random variable, observed as y, is assumed to follow a generalized linear model with a single covariate, x, that is measured with error and ordinarily other covariates, $\mathbf{u}$, that are observed without error. The relationship between the response, y, and these covariates will be written as $g(\varphi'\mathbf{w})$ where $g(\cdot)$ is a known function and

$$y_i \sim f_Y(\cdot;\mu_i)\ ;\ \mu_i = g(\varphi'\mathbf{w}_i) \text{ and } \varphi'\mathbf{w}_i = \alpha'\mathbf{u}_i + \beta x_i\ . \qquad 2.1.3\text{-}1$$

The case where $\mu_i = \varphi'\mathbf{w}_i$ is emphasized although the more general case where $\mu_i = g(\varphi'\mathbf{w}_i)$ is also discussed.

Within this class of regression models it will be adequate for our purposes to assume that $f_Y(\cdot;\mu)$ is a one (or two) parameter member of the exponential family. An advocated data analysis strategy exploits several properties of large cohort studies and, in these epidemiological applications, the usual regression models assume that the distribution of the response, $f_Y(\cdot;\mu)$, follows the binomial, exponential, normal, or Poisson distributions. These specific models are included in the class of distributions considered for $f_Y(\cdot;\mu)$. Survival models, although they are not explicitly considered here, are important and the principles to be developed also apply to them.

Many of the points here would be applicable for more general models for $f_Y(\cdot;\mu)$. The analytic findings largely depend upon a general score equation for μ given y and x. Thus conclusions concerning the role played by the distribution of the unobserved covariate, x, could undoubtedly be extended to other distributions or more complex regression relationships. However, it is better to restrict the class of models under consideration to generalized linear models in the exponential family in order to more simply address some of the factors that are of practical importance in epidemiological applications. Of importance in subsequent work is the equivalence, in the absence of measurement errors, of maximum likelihood estimates and iteratively reweighted least squares estimates when $f_Y(\cdot;\mu)$ is in the exponential family. Some

of the advantage that accompanies this equivalence would be lost under a more general formulation.

The reported simulations are restricted to a Poisson distribution for $f_Y(\cdot;\mu)$ and the examples are restricted to the Poisson and normal distributions. However, the actual estimation methods usually permit much more general choices for this distribution and they are initially presented in that context.

2.2 Structural Measurement Error Model

In this section, the maximum likelihood estimates and the iteratively reweighted least squares estimates are given for a structural model. A structural model for measurement error is constructed by specifying that the values of the unobserved covariate are a sample from a distribution. At least in principle, standard statistical theory is straightforward to apply in a structural model and, partly for that reason, this modeling approach is introduced first to establish some notation within a familiar context.

Suppose that the vector of unobserved covariates, $\mathbf{x}$, is an independent and identically distributed sample from the cumulative distribution, $F_X(\cdot)$. The distribution of y_i given the observed covariates, $\mathbf{u}_i$ and z_i ; $f_{Y|UZ}(\cdot)$, follows from this specification given the distributions, $f_{Z|X}(\cdot)$ and $f_Y(\cdot;\mu_i)$; $\mu_i = g(\varphi'\mathbf{w}_i)$. Denote the distribution of x given the observed covariates, $(\mathbf{u}_i, z_i)$; as

$$f_{X|UZ}(\mathrm{x}|\mathbf{u}_i, z_i; F_X)\,d\mathrm{x} \propto f_{Z|X}(z_i|\mathrm{x})\,dF_X(\mathrm{x}) , \qquad 2.2\text{-}1$$

then the distribution of y_i given the observed covariates is

$$f_{Y|UZ}(y_i|\mathbf{u}_i, z_i; F_X, \varphi) = \int f_Y(y_i; g(\varphi'\mathbf{w}_i))\, f_{X|UZ}(\mathrm{x}|\mathbf{u}_i, z_i; F_X)\,d\mathrm{x} . \qquad 2.2\text{-}2$$

This distribution (Equation 2.2-2) determines the likelihood of the observed data and is the foundation of likelihood methods. That is, the likelihood of the observed data is

$$\prod_{i=1}^{n} f_{Y|UZ}(y_i|\mathbf{u}_i, z_i; F_X, \varphi) . \qquad 2.2\text{-}3$$

The iteratively reweighted least squares estimator is of interest because it can be viewed as an approximation to the maximum likelihood estimator. Under the assumptions about $f_Y(\cdot;\mu)$ outlined in Section 2.1.3, the iteratively reweighted least squares estimator is equivalent to the maximum likelihood estimator in the absence of measurement errors and continuity considerations imply that it is an approximation at least for small measurement errors. There are several advantages in the use of iteratively reweighted least squares estimates that are elaborated in subsequent

sections. In the structural model, the mean and variance of the distribution, $f_{Y|UZ}(y_i|\mathbf{u}_i, z_i; F_X, \varphi)$, define the weighted least squares criterion,

$$\sum_{i=1}^{n} \frac{(y_i - E_{Y|UZ}[y])^2}{V_{Y|UZ}[y]} . \qquad 2.2\text{-}4$$

2.2.1 Maximum Likelihood Estimation

The likelihood equation for φ , $L(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z})$, can be expressed as

$$\begin{aligned} L(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) &= \prod_{i=1}^{n} \int \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X) dx \\ &= \prod_{i=1}^{n} E_{X|UZ}\left(\overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x)\right) \end{aligned} \qquad 2.2.1\text{-}1$$

where

$$\overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) = f_Y(y_i; g(\varphi' \mathbf{w})) . \qquad 2.2.1\text{-}2$$

That is, the likelihood for this structural model is simply the expected value, taken with respect to $f_{X|UZ}(\cdot)$, of the likelihood when there are no measurement errors.

In the absence of measurement error, $z_i = x_i$ and the naive likelihood,

$$\overset{*}{L}(\varphi; \mathbf{y}, \mathbf{u}, \mathbf{z}) = \prod_{i=1}^{n} \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, z_i) , \qquad 2.2.1\text{-}3$$

is the appropriate likelihood. The naive form will be indicated by an asterisk, $\overset{*}{L}$, throughout this thesis. Likewise, there are naive score equations and a naive information matrix. Expectations of these naive forms also appear in the corresponding equations of the structural model and an asterisk will also be used to distinguish them.

The score vector will be denoted as $U(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z})$ where the "r" th row of this vector is defined to be

$$\mathrm{U_r}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z}) = \sum_{i=1}^{n} \frac{\partial}{\partial \varphi_r} \ln\left[\int \overset{*}{\mathrm{L}}_i(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|UZ}(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right]. \qquad 2.2.1\text{-}4$$

In Appendix 1, Section A1-1.1, this score equation is shown to be equivalent to

$$\begin{aligned}\mathrm{U_r}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z}) &= \sum_{i=1}^{n} \int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx} \\ &= \sum_{i=1}^{n} \mathrm{E}_{X|YUZ=z_i}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\right].\end{aligned} \qquad 2.2.1\text{-}5$$

The expectation is with respect to the distribution of the unobserved x conditional on all the data, $(y_i,\mathbf{u}_i,z_i)$. This distribution is defined by the equation,

$$\mathrm{f}_{X|YUZ}(\mathrm{x}|y,\mathbf{u},z;\mathrm{F}_X) \propto \mathrm{f}_Y\left(y;\mathrm{g}(\boldsymbol{\alpha}'\mathbf{u}+\beta\mathrm{x})\right)\mathrm{f}_{X|UZ}(\mathrm{x}|\mathbf{u},z;\mathrm{F}_X). \qquad 2.2.1\text{-}6$$

Thus the score equations take the expectation of the naive score with respect to the conditional distribution of x given $(y_i,\mathbf{u}_i,z_i)$, $\mathrm{f}_{X|YUZ}(\cdot)$, in contrast to the likelihood, which depends on the conditional distribution of x given $(\mathbf{u}_i,z_i)$, $\mathrm{f}_{X|UZ}(\cdot)$. In iterative algorithms based on the score equations, this fact substantially increases computation since all integrals require reevaluation under the updated parameter estimates.

The observed information matrix, denoted by $\mathbf{i}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z})$, has the general term,

$$\begin{aligned}\mathrm{i_{rc}}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z}) &= \sum_{i=1}^{n} \int \overset{*}{\mathrm{i}}_{rci}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx} \\ &- \sum_{i=1}^{n} \int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \overset{*}{\mathrm{U}}_{ci}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx} \\ &+ \sum_{i=1}^{n}\left[\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx} \int \overset{*}{\mathrm{U}}_{ci}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\, \mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right]\end{aligned} \qquad 2.2.1\text{-}7$$

The derivation of this formula for the observed information is given in Appendix 1, Section A1-1.2.

The maximum likelihood estimate for this structural measurement error model, $\hat{\varphi}(F_X)$, will as a rule be a solution to $\mathbf{U}(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) = \mathbf{0}$. This equation can be solved by a Newton-Raphson iterative algorithm. The only problem beyond routine implementations of a Newton-Raphson algorithm is that the likelihood equations are determined as integrals over the range of x. In general, this will require a number of numerical integrations per iteration.

2.2.2 Iteratively Reweighted Least Squares Equations

The iteratively reweighted least squares solution, $\tilde{\varphi}(F_X)$, minimizes

$$\sum_{i=1}^{n} W_i\left(y_i - E_{Y|UZ}(y|\mathbf{u}_i, z_i)\right)^2 \qquad 2.2.2\text{-}1$$

where the weights, $W_i = V_{Y|UZ}^{-1}(y|\mathbf{u}_i, z_i)$, are evaluated using the current value for φ at each iteration of the fitting algorithm.

This procedure solution. Without measurement errors this procedure yields the maximum likelihood estimate in models from the exponential family (McCullagh and Nelder, 1989); in particular the models under discussion. So continuity considerations imply that this approximates the corresponding structural model maximum likelihood method at least for small covariate errors. When measurement errors are introduced, the iteratively reweighted least squares solution as outlined generally differs from the maximum likelihood solution. However, the iteratively reweighted least squares estimator is often relatively efficient in comparison to the maximum likelihood estimator. When a high relative efficiency is added in with the convenience and versatility of this method, iteratively reweighted least squares estimates become a very attractive alternative in measurement error problems. Noteworthy, regarding the relative efficiency of iteratively reweighted least squares estimates versus maximum likelihood estimates, is the model where the distributions, $F_X(\cdot)$, $f_Y(\cdot;\mu)$ and $f_{Z|X}(\cdot)$, are Gaussian distributions (Fuller, 1987; Kendall and Stuart, 1979). In this case, both methods yield the same estimator.

To apply an iteratively reweighted least squares algorithm, expressions for the mean and variance of y given $(\mathbf{u}_i, z_i)$ are required. In a structural model $f_{Y|UZ}(\cdot|\mathbf{u}_i, z_i; F_X, \varphi)$ is explicitly defined so that the needed expressions follow from their definitions. That is,

$$E_{Y|UZ}(y|\mathbf{u}_i, z_i) = \int y\, f_{Y|UZ}(y|\mathbf{u}_i, z_i; F_X, \varphi)\, dy \qquad 2.2.2\text{-}2$$

and

$$V_{Y|UZ}(y|\mathbf{u}_i, z_i) = \int (y - E(y|\mathbf{u}_i, z_i))^2\, f_{Y|UZ}(y|\mathbf{u}_i, z_i; F_X, \varphi)\, dy\ . \qquad 2.2.2\text{-}3$$

Under the distributional assumptions of this section, these reduce to

$$\begin{aligned} E_{Y|UZ}(y|\mathbf{u}_i, z_i) &= \int g(\alpha'\mathbf{u}_i + \beta x)\, f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx \\ &= E_{X|UZ}(g(\alpha'\mathbf{u}_i + \beta x)) \\ &= E_{X|UZ}(E_{Y|UX}(y)) \end{aligned} \qquad 2.2.2\text{-}4$$

and

$$\begin{aligned} V_{Y|UZ}(y|\mathbf{u}_i, z_i) &= \int V_{Y|UX=x}(y)\, f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx \\ &\quad + \int [g(\alpha'\mathbf{u}_i + \beta x)]^2\, f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx \\ &\quad - \left[\int g(\alpha'\mathbf{u}_i + \beta x)\, f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx\right]^2 \\ &= E_{X|UZ}(V_{Y|UX}(y)) + V_{X|UZ}(E_{Y|UX}(y))\ . \end{aligned} \qquad 2.2.2\text{-}5$$

Further simplification of these formulae occurs for a linear link, $g(\alpha'\mathbf{u} + \beta x) = \alpha'\mathbf{u} + \beta x$. Define

$$\xi_{F_X}(z) = \int x\, f_{X|UZ}(x|\mathbf{u}, z; F_X)\, dx = E_{X|UZ=z}(x) \qquad 2.2.2\text{-}6$$

and

$$V_{F_X}(z) = \int \left(x - \xi_{F_X}(z)\right)^2 f_{X|UZ}(x|\mathbf{u}, z; F_X)\,dx = E_{X|UZ=z}\left(x - \xi_{F_X}(z)\right)^2 , \qquad 2.2.2\text{-}7$$

then

$$E_{Y|UZ}(y|\mathbf{u}_i, z_i) = \alpha'\mathbf{u}_i + \beta\xi_{F_X}(z_i) \qquad 2.2.2\text{-}8$$

and

$$V_{Y|UZ}(y|\mathbf{u}_i, z_i) = \int V_{Y|UX=x}(y) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\,dx + \beta^2 V_{F_X}(z_i) . \qquad 2.2.2\text{-}9$$

In the special Gaussian distribution case many methods reproduce the maximum likelihood solution. Besides iteratively reweighted least squares, the correction for attenuation method (Section 2.5), the Nakamura method (Section 2.6), and the linear Schafer method (Section 2.7.2) yield this solution. A simple replacement; replacing z_i with $\xi_{F_X}(z_i)$ and proceeding as though there were no measurement errors, also yields the maximum likelihood estimates (Section 2.7.1). The application of replacement procedures to the generalized linear models being considered achieves, at the least, measurement error corrections that are computationally simple to implement. This simple correction does not yield the maximum likelihood estimates or even the iteratively reweighted least squares estimates, but in cases where the iteratively reweighted least squares procedure yields good approximations to the maximum likelihood estimator, the simpler replacement algorithm can be nearly as good. Further discussion of replacement methods and some factors that affect their efficiency relative to maximum likelihood estimates are taken up again in Section 2.7.

2.3 Functional Measurement Error Models

Functional models for measurement errors assume that the unobserved covariates, $\mathbf{x}$, are fixed constants. An estimation theory for the functional approach is not as developed as for structural models, and most of the theory that exists is based on the case where $f_Y(\cdot;\mu)$ and $f_{Z|X}(\cdot)$ are normal distributions (Amemiya and Fuller, 1988, Cochran, 1968; Kendall and Stuart, 1979; Fuller, 1987). Carroll *et al.* (1984) introduced a probit model for $f_Y(\cdot;\mu)$ but efforts to generalize to other cases leads to rather intractable mathematics (Stefanski and Carroll, 1985 and 1987).

A serious difficulty in advancing a general estimation method for functional models is that the unobserved values, $\mathbf{x}$, are incidental parameters in the resultant likelihood, i.e., nuisance parameters that are in 1-1 correspondence with the observations or small strata in the data. Estimation in situations with incidental parameters is often difficult, and the maximum likelihood estimates for φ are not necessarily consistent (Amemiya and Fuller, 1988, Fuller, 1987, and Stefanski and Carroll, 1985 and 1987). Thus standard statistical methods can not be relied upon in the functional setting. Conceptually the quasi-structural approach, which is essentially a structural approach that pertains to fixed $\mathbf{x}$, can avoid many of the problems and that model will be taken up in the next section.

It is well known that the naive estimates are biased and because standard theory often fails to find consistent estimators in the functional model, it becomes necessary to show that proposed estimators are unbiased for φ , at least asymptotically. To do this it is necessary to describe a sequence for the unobserved covariates, $(x)_n$. Hwang (1986) and Nakamura (1990) simply resort to a distribution, $F_X(\cdot)$, even though the underlying estimation method would usually be thought of as a functional solution. These models assume a general distribution for $\mathbf{x}$ which distinguishes them from more conventional structural models that assume a more specific distribution of $\mathbf{x}$. However, such subtleties blur the distinction between functional and structural models. Amemiya and Fuller (1988) avoid the outright assumption of a probability distribution, which generates the sequence, $(x)_n$, but the very assumption that a sequence exists seems tantamount to assuming such a distribution. The point is that the justifications for many functional methods rely on assumptions which as least come very near to implying a distribution for $\mathbf{x}$.

Stefanski and Carroll (1987) have developed a suitable theory for the functional model that maintains a purer perspective. The theory is built around unbiased score equations. However, the optimal estimator suggested by these criteria "will depend on the particular sequence of covariates", i.e. $F_X^n(\cdot)$. Consequently they tend to employ a structural setting to suggest estimators that satisfy the criteria that they have proposed. Nakamura (1990) gives a method for correcting score equations so that they are unbiased score equations in the sense of Stefanski and Carroll. More recent work by Stefanski and Carroll (1990 and 1991) has pursued efficient score tests based upon the unbiased score equations. Their work has a strong emphasis on methods to deconvolute $F_Z(\cdot)$ to obtain a nonparametric kernel density estimate for $E_{X|Z}\left(x;\hat{F}_X\right)$. The underlying philosophy is to avoid direct modeling of $F_X(\cdot)$ by estimating $\xi_G(z_i)$. This philosophy is closely aligned with empirical Bayes strategies such as the methods of Schafer (1990).

Two general approaches to the incidental parameters, $\mathbf{x}$, characterize most functional solutions. One is to replace the sufficient statistics defined through the naive likelihood,

$$\prod_{i=1}^{n} f_Y\left(y_i; g(\boldsymbol{\alpha}'\mathbf{u}_i + \beta x_i)\right), \qquad 2.3\text{-}1$$

with consistent estimates based upon $\mathbf{z}$ and $f_{Z|X}(\cdot)$. The other approach is to replace x_i in Equation 2.3-1 with an estimate of its expectation given z_i, such as $\xi_G(z_i)$. The correction for attenuation method (Section 2.5) and the Nakamura method (Section 2.6) are examples of the first approach and the Whittemore method (Section 2.7.1) and the Schafer method (Section 2.7.2) are examples of the second approach. One might consider another course as a third approach; simply adopting a structural model because of its mathematical convenience. This line of reasoning often justifies the use of structural models in practice. All three of these approaches can be viewed as approximate solutions to the quasi-structural score equations.

2.4 Quasi-structural Measurement Error Model

The quasi-structural model is primarily a theoretical tool that is used in this thesis to motivate and evaluate estimation methods. In the standard functional approach $\mathbf{x}$ is considered to be a vector of parameters. Pierce *et al.* (1992) have suggested that it may be more natural to think of $\mathrm{F}_X^n(\cdot)$ as the parameter, instead, and they define a quasi-structural likelihood for a functional setting, which depends on $\mathrm{F}_X^n(\cdot)$ rather than $\mathbf{x}$. In the context of this dissertation, this measurement error model is most directly defined as the structural model that results when the distribution of the unobserved covariate, $\mathrm{F}_X(\cdot)$, is replaced by the empirical distribution function, $\mathrm{F}_X^n(\cdot)$, in the equations of Section 2.2. The empirical distribution is defined whether or not $\mathbf{x}$ is considered to be a vector of parameters or a random sample from a distribution, $\mathrm{F}_X(\cdot)$. Consequently this model is defined in the functional setting and, at least in simulations, offers the mathematical stability of a structural model.

The empirical distribution, $\mathrm{F}_X^n(\cdot)$, is thought of as a nuisance parameter in the quasi-structural model. The maximum likelihood estimator, $\hat{\varphi}(\mathrm{F}_X^n)$, for this model considers $\mathrm{F}_X^n(\cdot)$ as known. Thus it is not an operational estimator but provides a useful point of reference for the performance of other estimators. A number of estimation methods solve equations that can be viewed as approximations to the quasi-structural score. The relative degree of approximation among these methods provides analytic insight into their efficiency. In particular, the two general approaches for dealing with the incidental parameters in traditional functional models make such an approximation. If the components of the quasi-structural score are written as

$$\mathrm{U}_{\mathrm{r}}(\varphi;\mathrm{F}_X^n,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\mathrm{E}_{X|YUZ=z_i}\left[\overset{*}{\mathrm{U}}_{ri}(\varphi;y_i,\mathbf{u}_i,\mathrm{x})\right], \qquad 2.4\text{-}1$$

then both approaches replace $\overset{*}{\mathrm{U}}_{ri}(\varphi;y_i,\mathbf{u}_i,\mathrm{x})$ with a function of $\mathbf{z}$. The correction for attenuation method (Fuller, 1987) and the Nakamura (1990) method make a substitution that gives consistent parameter estimates. Replacement methods, e.g.,

Whittemore (1989) and Schafer (1990), replace x_i with the expected value of x_i given z_i under an assumed distribution for x_i, $\xi_G(z_i)$.

In some cases, the nature of these approximations can be elucidated through a Taylor's expansion of the quasi-structural score. The quasi-structural score can be expanded about $\mathbf{x} = \boldsymbol{\gamma}$, to give

$$\sum_{i=1}^{n} \mathrm{E}_{X|\gamma UZ}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi}; y_i, \mathbf{u}_i, \mathrm{x})\right] \qquad 2.4\text{-}2$$

$$= \sum_{i=1}^{n}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi}; y_i, \mathbf{u}_i, \boldsymbol{\gamma}_i) + \sum_{k=1}^{\infty} \frac{\partial^k}{\partial \mathrm{x}^k} \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi}; y_i, \mathbf{u}_i, \boldsymbol{\gamma}_i) \frac{\mathrm{E}_{X|\gamma UZ}(\gamma_i - \mathrm{x})^k}{k!}\right].$$

As an approximation, correction for attenuation methods (Fuller, 1987) can be thought of as using $\boldsymbol{\gamma} = \mathbf{z}$. In principle, x^k in the terms, $(z_i - \mathrm{x})^k$, are replaced by $c_k z_i^k$ where $c_k \mathrm{E}_{Z|X=x}(z_i)^k = x^k$. Since the substituted quantity is not a function of x, the expectations, $\mathrm{E}_{X|\gamma UZ}(\cdot)$, do not depend upon the distribution, $\mathrm{F}_X^n(\cdot)$, or specification of a surrogate. As a result this approximation gives consistent estimates of $\boldsymbol{\varphi}$ (Nakamura, 1990). Section 2.6 elaborates on this "correction for attenuation" style of approximation in the context of its extension to corrected score equations, a method given by Nakamura (1990). Replacement methods, which are elaborated in Section 2.7, solve the score equations,

$$\sum_{i=1}^{n} \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi}; y_i, \mathbf{u}_i, \boldsymbol{\gamma}_i) = 0, \qquad 2.4\text{-}3$$

where γ_i is $\xi_G(z_i)$. In concept choices for $\boldsymbol{\gamma}$ which make the error,

$$\sum_{i=1}^{n}\sum_{k=1}^{\infty} \frac{\partial^k}{\partial \mathrm{x}^k} \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi}; y_i, \mathbf{u}_i, \boldsymbol{\gamma}_i) \frac{\mathrm{E}_{X|\gamma UZ}(\gamma_i - \mathrm{x})^k}{k!}, \qquad 2.4\text{-}4$$

small are best. Thus the accuracy of this approximation is directly related to the fidelity of $\mathrm{G}(\cdot)$ to $\mathrm{F}_X^n(\cdot)$.

2.5 Correction for Attenuation Method

In a classic normal theory measurement error model, the naive regression estimates can be corrected to the maximum likelihood estimates by a simple adjustment to the sum of squares, $\sum_{i=1}^{n}(z_i-\bar{z})^2$. In particular, assume that the variance of the measurement errors, σ^2 , is known and let

$$y|x \sim N(\alpha+\beta x,\sigma_y)\ ,\ z|x \sim N(x,\sigma) \text{ and } x \sim N(\mu,\sigma_x)\ , \qquad 2.5\text{-}1$$

where $\alpha,\beta,\mu,\sigma_y$ and σ_x are unknown parameters. Cochran (1968), Fuller (1987), and Kendall and Stuart (1979) discuss the derivation of the maximum likelihood estimates for this structural measurement error model. In this case,

$$\hat{\beta}=\frac{\sum_{i=1}^{n}y_i(z_i-\bar{z})}{\sum_{i=1}^{n}(z_i-\bar{z})^2-n\sigma^2} \qquad 2.5\text{-}2$$

whenever $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are nonnegative. This estimator is referred to as a correction for attenuation estimator since it increases the naive estimate so that it consistently estimates β . The attenuation of the naive estimator is corrected by the factor,

$$B=\frac{\sum_{i=1}^{n}(z_i-\bar{z})^2}{\sum_{i=1}^{n}(z_i-\bar{z})^2-n\sigma^2}\ . \qquad 2.5\text{-}3$$

The same correction for attenuation can be derived from a functional model. Assume that $z_i=x_i+e_i$ where x_i is fixed, $E(e_i)=0$ and $var(e_i)=\sigma^2$ then $E_{Z|X=x}(z)=x$ and $var_{Z|X=x}(z)=\sigma^2$. Under these assumptions, it is straightforward to show that

$$E_{Z|X}\left[\sum_{i=1}^{n}y_i(z_i-\bar{z})\right]=\sum_{i=1}^{n}y_i(x_i-\bar{x}) \qquad 2.5\text{-}4$$

and

$$\mathrm{E}_{Z|X}\left[\sum_{i=1}^{n}(z_i-\bar{z})^2\right]-n\sigma^2=\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad . \qquad 2.5\text{-}5$$

Thus, the estimator given by Equation 2.5-2 is a consistent estimator of β when $\mathbf{x}$ is assumed to be fixed.

This correction for attenuation estimator can be extended to models with known weights, W_i , and variances, σ_i^2 , to give

$$\hat{\beta}=\frac{\sum_{i=1}^{n}W_i y_i(z_i-\bar{z})}{\sum_{i=1}^{n}W_i(z_i-\bar{z})^2-\sum_{i=1}^{n}W_i\sigma_i^2} \quad , \qquad 2.5\text{-}6$$

which implies the adjustment,

$$B=\frac{\sum_{i=1}^{n}W_i(z_i-\bar{z})^2}{\sum_{i=1}^{n}W_i(z_i-\bar{z})^2-\sum_{i=1}^{n}W_i\sigma_i^2} \quad . \qquad 2.5\text{-}7$$

While the preceding formulation is based on additive errors, multiplicative errors can be incorporated by assuming that $z_i=x_i(1+e_i)$. Hwang (1986) discusses this model in a structural context. In a functional model where x_i is fixed, $\mathrm{E}(e_i)=0$ and $\mathrm{var}(e_i)=\nu^2$; $\mathrm{E}_{Z|X=x}(\mathrm{z})=x$ and $\mathrm{var}_{Z|X=x}(\mathrm{z})=\nu^2x^2$. From these relationships, it is not difficult to show that Equation 2.5-4 is satisfied,

$$\mathrm{E}_{Z|X}\left[\sum_{i=1}^{n}(z_i-\bar{z})^2-\sum_{i=1}^{n}\frac{\nu^2z_i^2}{\nu^2+1}\right]=\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad 2.5\text{-}8$$

and the attenuation of naive estimator is corrected by the factor,

$$B = \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2 - \sum_{i=1}^{n}\frac{\nu^2 z_i^2}{\nu^2 + 1}} . \qquad 2.5\text{-}9$$

These correction for attenuation estimators can also be viewed as replacement estimators. For an arbitrary $\kappa \in [0,1)$, consider a contraction of z_i toward $\bar{z}$ that is defined as

$$\xi_i = (1-\kappa)\bar{z} + \kappa z_i = \bar{z} + \kappa(z_i - \bar{z}) . \qquad 2.5\text{-}10$$

It follows immediately from the definition that the average, $\bar{\xi}$, is $\bar{z}$,

$$\sum_{i=1}^{n} W_i(\xi_i - \bar{\xi})^2 = \kappa^2 \sum_{i=1}^{n} W_i(z_i - \bar{z})^2 , \qquad 2.5\text{-}11$$

and

$$\sum_{i=1}^{n} W_i y_i(\xi_i - \bar{\xi}) = \kappa \sum_{i=1}^{n} W_i y_i(z_i - \bar{z}) . \qquad 2.5\text{-}12$$

The proposed replacement estimator would substitute ξ for $\mathbf{z}$ in the naive likelihood yielding the solution,

$$\frac{\sum_{i=1}^{n} W_i y_i(\xi_i - \bar{\xi})}{\sum_{i=1}^{n} W_i(\xi_i - \bar{\xi})^2} = \frac{\kappa \sum_{i=1}^{n} W_i y_i(z_i - \bar{z})}{\kappa^2 \sum_{i=1}^{n} W_i(z_i - \bar{z})^2} . \qquad 2.5\text{-}13$$

Consequently κ^{-1} is the implied correction to the naive slope estimate by the contraction, ξ, and taking $\kappa = B^{-1}$ gives $\hat{\beta}$, the maximum likelihood solution to the structural model defined by Equation 2.5-1.

Amemiya and Fuller (1988) and Hsiao (1989) discuss the extension of this structural model to nonlinear relationships, $g(\boldsymbol{\alpha}'\mathbf{u} + \beta x)$. Because the relationship between y and z will differ from the relationship between y and x, the theory

becomes cluttered by aspects related to the approximation of $g(\alpha'\mathbf{u}+\beta x)$ by $g(\alpha'\mathbf{u}+\beta z)$ and consequently the construction of consistent estimates is not so simple. For identity links, i.e., $g(\alpha'\mathbf{u}+\beta x)=\alpha'\mathbf{u}+\beta x$, in non Gaussian models for $f_Y(\cdot;\mu)$ such as the exponential family models considered by McCullagh and Nelder (1989), the naive estimates can be corrected by Equation 2.5-7. Consistency for β follows without much difficulty whenever the weights are constant (normal case) or consistently estimated with x replaced by z or a known function of z. Since the weights are approximately correct when evaluated at z, this adjustment of the naive estimator can be viewed as a simple approximation that corrects for attenuation. In the next section, a method that has been proposed by Nakamura (1990) will be elaborated in the context of the general model for $f_Y(\cdot;\mu)$. The Nakamura method correctly accounts for the effects of nonlinear links and weights that depend upon the regression estimates through an extension of the idea exemplified by Equations 2.5-5 and 2.5-8. In Section 2.7.1, the method of Whittemore is presented. Her method extents the correction for attenuation to general models for $f_Y(\cdot;\mu)$ by analogy with the replacement argument, i.e., Equation 2.5-10. While all three methods give estimates that coincide in the classic normal theory structural model that was discussed in this section, they are philosophically distinct and lead to different estimators in just about any other circumstance.

2.6 Nakamura Method

Nakamura (1990) advanced an estimation method in which the naive score equations are corrected to give consistent parameter estimates whenever the conditional distribution of the observed covariate, $f_{Z|X}(\cdot)$, is known. The method is based on a corrected score, $\dot{\mathbf{U}}^{C}(\varphi;\mathbf{y},\mathbf{u},\mathbf{z})$, which adjusts the naive score, for example,

$$\dot{\mathbf{U}}^{C}(\varphi;\mathbf{y},\mathbf{u},\mathbf{z}) = \dot{\mathbf{U}}(\varphi;\mathbf{y},\mathbf{u},\mathbf{z}) - \mathbf{t}(\mathbf{z};\varphi), \qquad 2.6\text{-}1$$

where $\mathbf{t}(\mathbf{z};\varphi)$ is chosen to satisfy the relationship,

$$\mathrm{E}_{Z}\left[\dot{\mathbf{U}}^{C}(\varphi;\mathbf{y},\mathbf{u},\mathbf{z})\right] = \mathrm{E}_{X}\left[\dot{\mathbf{U}}(\varphi;\mathbf{y},\mathbf{u},\mathbf{x})\right]. \qquad 2.6\text{-}2$$

Nakamura demonstrated that corrected score equations, i.e., equations that satisfy Equation 2.6-2, give consistent parameter estimates. These estimates will be denoted by $\hat{\varphi}$. Sometimes it may not be possible to find a corrected score equation. For example, this is the case for a Poisson model with multiplicative errors that is discussed further in Section 4.5. In general, this definition can be relaxed so that Equation 2.6-2 holds asymptotically and the corrected score will yield consistent estimates. In this work, this relaxed definition will be used whenever corrected score equations cannot be found using the more restrictive version.

If the expectation of the naive score is written as $\dot{\mathbf{U}}(\varphi;\mathbf{y},\mathbf{u},\mathbf{x})$ plus whatever remains, i.e.,

$$\mathrm{E}_{Z|X}\left[\dot{\mathbf{U}}(\varphi;\mathbf{y},\mathbf{u},\mathbf{z})\right] = \dot{\mathbf{U}}(\varphi;\mathbf{y},\mathbf{u},\mathbf{x}) + \mathbf{T}(\mathbf{x};\varphi), \qquad 2.6\text{-}3$$

then the function, $\mathbf{t}(\mathbf{z};\varphi)$, needed to satisfy Equation 2.6-2 can be any function with expectation, $\mathrm{E}_{Z|X}\left[\mathbf{t}(\mathbf{z};\varphi)\right]$, that equals $\mathbf{T}(\mathbf{x};\varphi)$. Or, to relax the requirement so that Equation 2.6-2 holds asymptotically, $\mathbf{t}(\mathbf{z};\varphi)$ can be any function that converges to

$\mathbf{T}(\mathbf{x};\boldsymbol{\varphi})$. The remainder, $\mathbf{T}(\mathbf{x};\boldsymbol{\varphi})$, can be found by expanding the naive score equations in a Taylor's series about $z_i = x_i$ to obtain the equation,

$$\mathrm{E}_{Z|X}\left[\sum_{i=1}^{n} \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,z_i)\right] \qquad 2.6\text{-}4$$
$$= \sum_{i=1}^{n}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,x_i) + \sum_{k=1}^{\infty}\frac{\partial^k}{\partial x_i^k}\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,x_i)\frac{\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^k}{k!}\right].$$

Since the assumptions of Section 2.1.2 imply that $\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)=0$ and $\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^k$ is a known function of x_i, the score equations defined in Equation 2.6-4 can be rewritten as

$$\mathrm{E}_{Z|X}\left[\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})\right] = \overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{x}) + \sum_{k=2}^{\infty}\mathbf{T}_k(\mathbf{x},\boldsymbol{\varphi}) \qquad 2.6\text{-}5$$

where

$$\mathbf{T}_k(\mathbf{x};\boldsymbol{\varphi}) = \begin{bmatrix} \sum_{i=1}^{n}\frac{\partial^k}{\partial x_i^k}\overset{*}{\mathrm{U}}_{1i}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,x_i)\frac{\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^k}{k!} \\ \vdots \\ \sum_{i=1}^{n}\frac{\partial^k}{\partial x_i^k}\overset{*}{\mathrm{U}}_{pi}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,x_i)\frac{\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^k}{k!} \end{bmatrix}. \qquad 2.6\text{-}6$$

The fundamental idea then is to replace each $\mathbf{T}_k(\mathbf{x};\boldsymbol{\varphi})$ with a function of $\mathbf{z}$, $\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi})$, such that

$$\lim_{n\to\infty}\left|\mathrm{E}_{Z|X}\left[\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi})\right]-\mathbf{T}_k(\mathbf{x};\boldsymbol{\varphi})\right|=0\ . \qquad 2.6\text{-}7$$

Then the corrected score can be defined as

$$\overset{*\,\mathrm{C}}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})=\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})-\sum_{k=2}^{\infty}\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi}) \qquad 2.6\text{-}8$$

In many practical situations this sum is finite (see Appendix 1, Section A1-2) or can be truncated (see Section 4.5). From this definition it follows that

$$\begin{aligned}\mathrm{E}_Z\left[\overset{*\,\mathrm{C}}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})\right]&=\mathrm{E}_X\left[\mathrm{E}_{Z|X}\left[\overset{*\,\mathrm{C}}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})\right]\right]\\&=\mathrm{E}_X\left[\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{x})+\sum_{k=2}^{\infty}\mathbf{T}_k(\mathbf{x},\boldsymbol{\varphi})-\mathrm{E}_{Z|X}\left[\sum_{k=2}^{\infty}\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi})\right]\right] \qquad 2.6\text{-}9\\&=\mathrm{E}_X\left[\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{x})\right]+\sum_{k=2}^{\infty}\mathrm{E}_X\left[\mathbf{T}_k(\mathbf{x},\boldsymbol{\varphi})-\mathrm{E}_{Z|X}\left[\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi})\right]\right]\end{aligned}$$

and this gives

$$\lim_{n\to\infty}\left|\mathrm{E}_Z\left[\overset{*\,\mathrm{C}}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{z})\right]-\mathrm{E}_X\left[\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\mathbf{x})\right]\right|=0\ . \qquad 2.6\text{-}10$$

The Nakamura method is considered to be a generalization of correction for attenuation methods primarily because $\mathbf{t}_k(\mathbf{z};\boldsymbol{\varphi})$ is found through method of moments arguments along the lines discussed in Section 2.4 and exemplified in Equations 2.5-7, 2.5-8, and 2.5-11. The simple corrections for the naive estimates, which have been outlined in Section 2.5, are recovered by the method and this is illustrated for three measurement error models in Appendix 1, Section A1-2. However, Nakamura developed the method to generate consistent parameter estimates under an arbitrary

distribution for $F_X(\cdot)$, so the functional - structural distinction is not clearly made. As an extension of correction for attenuation concepts, the method is algebraically more tractable when the solution to the naive score equations cannot be expressed in closed form. Thus this method is more generally applicable to the class of models under discussion.

In Section 4.5, a procedure based on the Nakamura method is proposed for a Poisson regression model with multiplicative errors. In this case, the derivatives of the naive score equations are also functions of $\mathbf{x}$ and this further complicates finding a correction. Still, the procedure that was outlined above with the Taylor's expansion truncated at the second term gave an approximation to the Nakamura correction that turned out to be essentially unbiased in the simulations. Recently, Nakamura (1992) presented a similar approximation for a proportional hazards model when the covariate errors are additive. In these cases, $\mathbf{T}_k(\mathbf{x};\varphi)$ can be expressed as a ratio of polynomials in $\mathbf{x}$ and replacing x_i^k with $c_k z_i^k$ where $c_k \mathrm{E}_{Z|X=x_i}(\mathrm{z}^k) = x_i^k$ gives a suitable $\mathbf{t}_k(\mathbf{z};\varphi)$.

A Nakamura corrected score can also be viewed as an approximation to the quasi-structural score equations, Equation 2.4-1. Formally replacing $\dot{\mathrm{U}}_{ri}(\varphi; y_i, \mathbf{u}_i, \mathrm{x})$ with the corrected score, $\dot{\mathrm{U}}^{\mathrm{C}}_{ri}(\varphi; y_i, \mathbf{u}_i, z_i)$, in the quasi-structural score equations indicates that the solution to the corrected score equations is an approximate solution of the quasi-structural score equations. Nakamura (1990) exploits this substitution to prove that the corrected equations will consistently estimate the regression parameters for arbitrary $F_X(\cdot)$.

Our interest in the connection to the quasi-structural model will be in contrasting the degree of approximation implied by these estimates to the approximation implied by other methods. Through such a comparison the relative merits of estimators can be deduced. For this purpose, it is instructive to look at the step given in Equation 2.6-4 as it relates to the quasi-structural score equation expansion, Equation 2.4-2. In essence, the expectations, $\mathrm{E}_{X|Z}(z_i - \mathrm{x})^k$, in the quasi-structural score expansion are replaced with functions of $\mathbf{z}$ that are consistent for the expectations, $\mathrm{E}_{Z|X=x_i}(\mathrm{z} - x_i)^k$. This approximation is cruder than $\mathrm{E}_{X|Z}(z_i - \mathrm{x})^k$

which is the basic substitution of replacement methods. Since replacement methods truncate the expansion and the Nakamura method can involve other approximations as well, the overall comparison is more complex but the perception is that a good replacement method would yield the better approximation.

2.7 Replacement Methods

Replacement methods simply replace the unobserved value, x_i, in the naive likelihood with an estimate of the expected value, $\mathrm{E}_{X|Z=z_i}(\mathrm{x})$, and then proceed with the analysis as if x_i were observed. That is, a replacement method solves the score equation,

$$\overset{*}{\mathbf{U}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\xi_{\mathrm{G}}(\mathbf{z}))=\mathbf{0}\ , \tag{2.7-1}$$

to obtain an estimate, $\hat{\overset{*}{\boldsymbol{\varphi}}}(\xi_{\mathrm{G}})$, where $\xi_{\mathrm{G}}(\mathbf{z})$ is defined by

$$\xi_{\mathrm{G}}(z_i)=\int \mathrm{x}\,\mathrm{f}_{X|UZ}(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{G})\,\mathrm{dx}=\mathrm{E}_{X|UZ}(\mathrm{x})\ . \tag{2.7-2}$$

At least when measurement errors are small, the replacement estimator closely approximates the maximum likelihood estimator.

A Taylor's expansion of the naive log-likelihood about $\mathbf{x}=\xi_{\mathrm{G}}(\mathbf{z})$ yields an expression for the error of the approximation;

$$\begin{aligned}&\ln \mathrm{L}(\boldsymbol{\varphi};\mathrm{G},\mathbf{y},\mathbf{u},\mathbf{z})-\ln \overset{*}{\mathrm{L}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\xi_{\mathrm{G}}(\mathbf{z}))\\ &\approx\sum_{i=1}^{n}\ln\left[\mathrm{E}_{X|UZ}\left[\exp\left(\frac{\partial \ln \overset{*}{\mathrm{L}}(\boldsymbol{\varphi};\mathbf{y},\mathbf{u},\xi_{\mathrm{G}}(\mathbf{z}))}{\partial\mu}\frac{\partial\mu}{\partial\mathrm{x}}(\mathrm{x}-\xi)\right)\right]\right].\end{aligned} \tag{2.7-3}$$

This expression is derived in Appendix 1, Section A1-3. For an identity link, i.e. $\mu = \boldsymbol{\alpha}'\mathbf{u}+\beta x$, the magnitude of this difference between these log-likelihoods is on the order of

$$\sum_{i=1}^{n}\frac{\beta^2 V_{\mathrm{G}}(z_i)}{\mathrm{V}_{Y|UX=\xi_{\mathrm{G}}(z_i)}(\mathrm{y})} \tag{2.7-4}$$

where the variance, $V_{Y|UX=\xi_G(z_i)}(y)$, is the variance of y under $f_Y(\cdot;\mu)$ evaluated at $\mathbf{u}_i$ and $x=\xi_G(z_i)$. The expression is further complicated by the nonlinear form of the general link, $g(\alpha'\mathbf{u}+\beta x)$, but the difference, Equation 2.7-3, remains on the order of the expression in Equation 2.7-4.

Within the class of models for $f_Y(\cdot;\mu)$ that are being considered, the score equations, Equation 2.7-1, can be solved by an iteratively reweighted least squares algorithm. Then, for an identity link, $g(\alpha'\mathbf{u}+\beta x) = \alpha'\mathbf{u}+\beta x$, direct application of the replacement method closely duplicates the weighted least squares solution whenever $\beta^2 V_G(z_i)$ is negligible. A practical advantage of the iteratively reweighted least squares formulation is that the component, $\beta^2 V_G(z_i)$, is added to the weight. This effectively adds a component of variance that eliminates much of the bias from estimated standard errors for the regression estimators. Otherwise, the standard errors tend to be underestimated. In the situations of most interest, the iteratively reweighted least squares regression estimates essentially duplicate the solution to Equation 2.7-1 so the primary practical difference between the methods is in the estimated standard errors.

The replacement approach is not limited to an identity link. Equation 2.7-1 places no restriction on the link function. Likewise, the iteratively reweighted least squares model of Section 2.3 certainly does not require linearity although the general form makes additional approximations. The crux of these additions lies largely with the approximation,

$$g(\alpha'\mathbf{u}+\beta\xi_G(z)) \approx \int g(\alpha'\mathbf{u}+\beta x)\, f_{X|UZ}(x|\mathbf{u},z;G)\,dx \qquad 2.7\text{-}5$$

A Taylor's expansion of this integral about $\xi_G(\mathbf{z})$ gives the difference,

$$\left|E_{X|UZ}[g(\alpha'\mathbf{u}+\beta x)] - g(\alpha'\mathbf{u}+\beta\xi_G(z))\right| \approx \beta^2 V_G(z)\frac{\partial^2}{\partial x^2}[g(\alpha'\mathbf{u}+\beta\xi_G(z))] \qquad 2.7\text{-}6$$

Once again, $\beta^2 V_G(z_i)$ being small is conducive to good approximations. Of more general importance, the local curvature affects the approximation and most link functions are smooth curves. Looking directly at Equation 2.7-6, it is clear that if

$g(\alpha' \mathbf{u} + \beta x)$ can be approximated well by a line on the main support for $f_{X|UZ}(x|\mathbf{u}, z; G)$ then the approximation will be very good.

The replacement approach is attractive for several reasons. An obvious attraction is that this approach usually avoids much of the computational difficulty often associated with the structural model solution. Even more important than simpler calculations, this approach separates the estimation algorithm into two steps, (1) correcting for the errors in the covariate and (2) estimating the regression. Separation of the estimation problem into simpler steps aids understanding and presenting the results especially in cases where the assumptions required for a solution need verification. Estimating $\xi_G(\mathbf{z})$ is a classic empirical Bayes problem and some of the estimation procedures that are available are more robust in their assumptions about $G(\cdot)$ than is generally true of structural models (Maritz and Lwin, 1989). This may allow one to orient away from the specific distributional assumptions and to focus on the "scatter plot" properties of the underlying distributions. That should facilitate an analysis strategy that results in estimates that are more robust to specific distributional assumptions. The method of Schafer (1990, 1992) is a good example of this approach.

Replacement methods can also be considered as an approximation to a quasi-structural score equation by simply replacing γ with $\xi_G(\mathbf{z})$ in Equation 2.4-2. When $G(\cdot)$ closely approximates $F_X^n(\cdot)$, the replacement estimator can be very efficient relative to the quasi-structural maximum likelihood estimator. The view adopted here is that the replacement method is approximating the quasi-structural likelihood with the first term of its Taylor's expansion about $\mathbf{x} = \xi_G(\mathbf{z})$.

This substitution into the Taylor's expansion implies an error due to the difference, $E_{X|YUZ}(x) - \xi_G(z)$. From this result it would seem that $E_{X|YUZ}(x)$ would be a better choice for γ_i. However, because $E_{X|YUZ}(x)$ depends upon φ, this choice greatly diminishes the computational convenience. In general, to capture the computational convenience afforded by weighted least squares using adjustments to $\mathbf{z}$ one must dispense with the information about $\mathbf{x}$ contained in $\mathbf{y}$.

There seem to be additional problems with estimates for $E_{X|YUZ}(x)$ that may well preclude their use even if it was computationally convenient. Since these

problems are poorly understood this discussion will be limited to some conjectures and observations. In the standard normal theory errors-in-variable model, i.e. the model defined in Equations 2.5-1, $\xi_G(\mathbf{z})$ in a replacement model yields the maximum likelihood solution. Further, in some simulation experiments, $E_{X|YUZ}(x)$ performed poorly except in those situations where $E_{X|YUZ}(x) \approx \xi_G(z)$. This problem extends to simulations under a Poisson model for $f_Y(\cdot;\mu)$. The conjecture is that identifiability problems attend the use of the information in $\mathbf{y}$. In more naive terms, $\mathbf{y}$ can be used to estimate β if $\mathbf{x}$ is given or $\mathbf{x}$ if β is given but it can't simultaneously estimate both quantities. Procedures that attempt to get around the basic lack of identifiability by an algorithm that conditions on β to estimate $E_{X|YUZ}(x)$ and then conditions on $E_{X|YUZ}(x)$ to update β do not work.

The efficiency of the replacement estimator relative to the maximum likelihood estimator will be highest when $\beta^2 V_G(z_i)/V_{Y|UX=\xi}(y|\mathbf{u}_i, \xi_G(z_i))$ is small. This condition implies that there is little information about $\mathbf{x}$ to be gained from $\mathbf{y}$. In cohort studies, this condition often is satisfied so that the most important aspect of the search for a good estimate depends upon specifying a distribution to approximate $F_X^n(\cdot)$. Two methods, one due to Whittemore (1989) and the other due to Schafer (1990) will be elaborated to exemplify the method and to illustrate the importance of choosing $G(\cdot)$.

2.7.1 Whittemore Method

Whittemore (1989) presented a replacement method that applies to additive covariate errors. In particular, let $z = x + \varepsilon$ where σ^2 is known and

$$\begin{bmatrix} x \\ \varepsilon \end{bmatrix} \sim \text{MVN}\left[\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}\right] \qquad 2.7.1\text{-}1$$

Then

$$\xi_\Phi(z_i) = (1-\kappa)\bar{z} + \kappa z_i \qquad 2.7.1\text{-}2$$

and

$$V_{\Phi}(z_i) = \kappa\sigma^2 \qquad 2.7.1\text{-}3$$

where $\kappa = B^{-1}$ and B is given in Equation 2.5-7. Using a weighted least squares formulation that incorporates the over dispersion induced by $V_{\Phi}(\mathbf{z})$ improves the estimated variance of the regression parameters.

This method can be modified to lognormal distributions (i.e. multiplicative errors) by applying the adjustment to the logarithms. To see this, let $z = x \times \varepsilon$, then $\ln(z) = \ln(x) + \ln(\varepsilon)$. Take

$$\begin{bmatrix} \ln(x) \\ \ln(\varepsilon) \end{bmatrix} \sim \text{MVN}\left[\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}\right] \qquad 2.7.1\text{-}4$$

then, for $w_i = \ln(z_i)$, $\xi_{\Phi}(w_i) = (1-\kappa)\overline{w} + \kappa w_i$ and $V_{\Phi}(w_i) = \kappa\sigma^2$ where

$$\kappa = \frac{\sum_{i=1}^{n}(w_i - \overline{w})^2 - n\sigma^2}{\sum_{i=1}^{n}(w_i - \overline{w})^2} \qquad 2.7.1\text{-}5$$

That is, the distribution of $\ln(x)$ conditional on w is a normal distribution with expectation, $\xi_{\Phi}(w)$, and variance, $V_{\Phi}(w)$. Consequently, $f_{X|Z}(\cdot|z;\mathrm{G})$ is a lognormal distribution with these parameters and it has expectation,

$$\xi_{\mathrm{G}}(z_i) = \exp\left(\xi_{\Phi}(w_i) + \frac{V_{\Phi}(w_i)}{2}\right), \qquad 2.7.1\text{-}6$$

and variance,

$$V_{\mathrm{G}}(z_i) = \exp\left(2\xi_{\Phi}(w_i) + 2V_{\Phi}(w_i)\right) - \left[\xi_{\mathrm{G}}(z_i)\right]^2 . \qquad 2.7.1\text{-}7$$

The only difficulty with this development is the assumption in Equation 2.7.1-4 that $E[\ln(\varepsilon)] = 0$, which implies that $E[\varepsilon] \neq 1$ and, consequently, $E_{Z|X=x_i}[z] \neq x_i$. To comply with the assumption, $E_{Z|X=x_i}[z] = x_i$, of Section 2.1.2, $f_{Z|X}(\cdot)$ should be a lognormal density satisfying this expectation. Let $\dot{w}_i = w_i - \sigma^2/2$ so that $E_{Z|X=x_i}[z] = x_i$. Then it is straightforward to demonstrate that $\xi_\Phi(\dot{w})$ and $V_\Phi(\dot{w})$ are algebraically equal to $\xi_\Phi(w)$ and $V_\Phi(w)$. Hence, for computational purposes it is satisfactory to work with the distributional assumptions as given in Equation 2.7.1-4.

2.7.2 Schafer Method

An estimation procedure, which is believed to be more adaptable to various shapes in $F_X(\cdot)$, was proposed by Schafer (1990). This method is a replacement method that utilizes a polynomial empirical Bayes estimator for $\xi_G(z)$ (Maritz and Lwin, 1989). Consider using a polynomial in z,

$$\xi_q(z) = \sum_{s=0}^{q} \theta_s z^s \qquad 2.7.2\text{-}1$$

as an estimate for x where the vector of coefficients, θ, is chosen to minimize the expected squared error,

$$\iint \left[\xi_q(z) - x\right]^2 f_{Z|X}(z)\,dz\,dG(x) \qquad 2.7.2\text{-}2$$

Taking derivatives with respect to θ generates the system of equations,

$$\frac{\partial}{\partial \theta_r}\left[\iint \left[\xi_q(z) - x\right]^2 f_{Z|X}(z)\,dz\,dG(x)\right] = 0\,, \qquad 2.7.2\text{-}3$$

which reduces to

$$\sum_{s=0}^{q} \theta_s E_Z(z^{s+r}) = \int x\, E_{Z|X}(z^r)\, dG(x)\ ;\ r = 0,\ 1,\ \cdots\ q\ . \qquad 2.7.2\text{-}4$$

To illustrate a solution to Equations 2.7.2-4 take $q = 1$ under a multiplicative error model for $f_{Z|X}(\cdot)$ with constant coefficient of variation, ν, then $E_{Z|X=x}(z) = x$ and $E_{Z|X=x}(z^2) = x^2(\nu^2 + 1)$ which gives the equations,

$$\theta_0 + \theta_1 E_Z(z) = E_X(x) \text{ and } \theta_0 E_Z(z) + \theta_1 E_Z(z^2) = E_X(x^2). \qquad 2.7.2\text{-}5$$

Noting that $E_Z(z) = E_X(x)$ and $\frac{E_Z(z^2)}{\nu^2 + 1} = E_X(x^2)$ yields the linear empirical Bayes estimates, $\xi_1(z) = \hat{\theta}_0 + \hat{\theta}_1 z$, where $(\hat{\theta}_0, \hat{\theta}_1)$ satisfy

$$\hat{\theta}_0 + \hat{\theta}_1 \bar{z} = \bar{z} \text{ and } \hat{\theta}_0 \bar{z} + \frac{\hat{\theta}_1 \sum_{i=1}^{n} z_i^2}{n} = \frac{\sum_{i=1}^{n} z_i^2}{n(\nu^2 + 1)}. \qquad 2.7.2\text{-}6$$

This is the linear estimator of Schafer (1990) for a multiplicative error model.

Similar solutions are available for additive error models. The linear estimator in the additive case is identical to that of the Whittemore method (Equation 2.7.1-1). Thus a quadratic empirical Bayes estimator $(q = 2)$ is more adaptable than the linear empirical Bayes estimator (Schafer, 1990). The quadratic equations used in the simulations are developed in Section 4.6.2.

Chapter 3. Modeling Covariate Measurement Error in Cohort Studies

The main conjectures of this dissertation, as listed in Chapter 1, are:

1. It is important that methods explicitly and carefully consider the distribution of the true covariate values. In the functional setting, the empirical distribution can play this role through the quasi-structural model.
2. This point is far more important in large samples than in small ones, and hence critical in epidemiological studies, which typically involve very large samples.
3. It is inadequate to focus only on how well an estimator removes the bias present in the naive estimates. There will be various methods which adequately remove this bias, but they may differ greatly in their "efficiency"; that is, how closely they come to the estimates based upon knowing the true covariate values.
4. In particular as the sample size increases, it seems plausible that the classical "correction for attenuation" approach will be increasingly less efficient relative to the best possible solution.

The primary aim, in this chapter, is to present some simulation results and some conclusions that pertain to these points. The exposition of the simulation results and conclusions is preceded by a few sections which are intended to elaborate on some ideas that support the choices for the measurement error methods and the scenarios that were simulated. First there is a discussion of some cohort study attributes that are relevant to approaches that might be taken in modeling measurement errors. This is followed by a more specific discussion of the particular characteristics from a cohort study, the Life Span Study, that were emulated in these simulations. The simulations compare a measure of "relative efficiency" that is also briefly motivated and explained. Except for a brief sketch, the more technical aspects of the simulation methods are deferred to Chapter 4.

3.1 Characteristics of Cohort Studies

The main postulate of this work is that efficient estimation in measurement error models requires attention to the distribution of the true covariate values. In considering the implementation of this postulate, three aspects of cohort studies contribute to make replacement methods attractive. They are:

1. large samples and non Gaussian distributions; a combination that makes modeling the distribution of the true covariate values a major factor in establishing efficiency.
2. responses, y, that are not very informative about the unobserved x; a situation that makes replacement and iteratively reweighted least squares methods nearly as efficient as the corresponding maximum likelihood methods.
3. analyses that are exploratory in nature; a purpose favoring transparent methods, versatile methods, and methods that are easy to use.

Under these conditions replacement methods can be efficient compared to other approaches as well as being relatively transparent, versatile, and easy to implement. That is, a replacement method stemming from a surrogate distribution for $\mathrm{F}_X^n(\cdot)$ should be relatively efficient. This distribution presumably would be selected to mimic the salient properties of the empirical distribution as they are evidenced by the observed covariate.

A large sample furnishes an abundance of information on the distribution of the true covariates and, especially when $\mathrm{f}_{Z|X}(\cdot)$ is assumed to be known, much about the form of $\mathrm{F}_X^n(\cdot)$ can be inferred from the observed empirical evidence provided by $\mathbf{z}$. In these situations, an *ad hoc* choice for $\mathrm{G}(\cdot)$ that mirrors the empirical evidence would seem preferable to reliance on a traditional method that only crudely approximates $\mathrm{F}_X^n(\cdot)$.

It is necessary to consider the efficiency of methods that avoid explicit specification of $\mathrm{F}_X(\cdot)$, such as correction for attenuation methods or their generalization, the Nakamura method. Algebraically, these methods are often nearly equivalent to a replacement method based upon a Gaussian choice for $\mathrm{F}_X(\cdot)$. In

these cases, their efficiency can be little better than the "equivalent" replacement method.

More generally, the efficiency of the quasi-structural replacement estimator, $\hat{\hat{\varphi}}\left(\xi_{F_X^n}\right)$, relative to that of the Nakamura estimator, $\hat{\varphi}$, can be considered through the approximation to the quasi-structural score that is implied by their respective estimating equations. Essentially $\hat{\hat{\varphi}}\left(\xi_{F_X^n}\right)$ approximates $E_{X|YUZ}(\cdot)$ with $E_{X|UZ}(\cdot)$ while the Nakamura estimator uses a method of moments approximation based upon $E_Z(\cdot) = E_X\left[E_{Z|X}(\cdot)\right]$. The relative accuracy of these approximations implies that quite generally the "correction for attenuation" methods will be less efficient than an optimum replacement method.

Some empirical Bayes estimators, also, provide estimates, ξ, that are perceived to be more adaptable to various forms of $F_X^n(\cdot)$. An example is the Schafer estimator, a polynomial empirical Bayes estimator. Although the two approaches to estimation are fundamentally different, the Schafer estimator and the Nakamura estimator take advantage of $E_{Z|X}(\cdot)$ in basically the same way. For this reason it seems likely that the Schafer estimator will have comparable efficiency to the Nakamura method.

In large cohorts where the sample sizes run to several thousands it is feasible to approximate $F_X^n(\cdot)$ with a deconvolution of $F_Z^n(\cdot)$ using the knowledge of $f_{Z|X}(\cdot)$ (Pierce *et al.*, 1992). This approach or a similar approach that deliberately uses the abundant information in $F_Z^n(\cdot)$ seems most likely to produce the more efficient replacement estimators.

The strategy does not categorically preclude traditional methods but it does prescribe a close approximation to $F_X^n(\cdot)$. The frequency distribution of true exposures can have an arbitrary pattern and as a result most specific methods are simply too inflexible to be universally applied in cohort studies. Because such methods are somewhat limited in their ability to adapt to a range of shapes, a suitable implementation of a method would frequently require some innovation in this aspect of the measurement error model. Since $\xi_G(z)$ is a simpler function than the

likelihood, such innovation is inherently easier to incorporate as a replacement method.

Other relevant aspects of cohort studies have to do with attributes that make the efficiency of a replacement method competitive with the corresponding maximum likelihood estimator. In this regard, the most important trait of cohort data is that the response, "case" or "not a case", is not very informative about the level of an exposure. In this situation the approximation, $\mathrm{E}_{X|YZ}(\cdot) \approx \mathrm{E}_{X|Z}(\cdot)$, will be very good and an iteratively reweighted least squares / replacement estimator will be nearly as efficient as the corresponding maximum likelihood estimator. Although it is not as important a characteristic, the frequently shallow dose response further enhances the efficiency of the replacement estimator.

A more obvious advantage of replacement methods is their versatility and in circumstances where they are also relatively efficient at estimating φ this makes them very attractive. Versatility is quite important in cohort analyses. The mortality rates or morbidity rates are known to depend upon age, sex and other factors. These effects must be fit to parsimoniously model any alterations in the rates that can be attributed to the exposure. A number of causes of mortality or morbidity are ordinarily monitored and the regression models for each "cause" typically evolve through a series of exploratory analyses.

Most of the points discussed above can be demonstrated in simulations. It should be clear from the discussion that several characteristics of the data affect the relative efficiency among methods and the course adopted here was to mimic conditions in a particular study, the Life Span Study. A description of this study as it pertains to the simulations follows.

3.1.1 Example: Life Span Study of Atomic Bomb Survivors

The Life Span Study (LSS) is a cohort study that has been analyzed by the replacement approach (Pierce *et al.*, 1992). The LSS follows the causes of mortality in a cohort of 76,000 persons, about half being controls and the remainder survivors of the atomic bombings of Hiroshima and Nagasaki. The frequency distribution of exposures among the survivors is dominated by lower exposures; partially because the cohort is composed of survivors but principally because a much larger initial population received the lower exposures. That is, the frequency of the true exposures

declines rapidly with increasing exposure, a very non Gaussian distribution. In this case, Pierce *et al.* (1990, 1992) employed a Weibull distribution based upon the perceived form of the true distribution and selected parameters to be consistent with the empirical distribution of the observed covariate, $F_Z^n(\cdot)$.

The LSS illustrates another "versatility" aspect of replacement methods. An essential undertaking was to ascertain a radiation exposure for each member of the cohort. The natural course was to divide the overall problem of estimating a dose-response into two tasks; estimate the dose and, given the dose, estimate the dose-response. This division occurred because much of the dosimetry involved experiments and sophisticated assessments of the transmission of various energies and types of radiation through shielding materials. Chiefly work by Jablon (1971) and Gilbert (1982) has demonstrated that a lognormal model with constant coefficient of variation is a credible probability model for these errors. Given this development, replacement methods are an extension of the overall process of assessing exposures. While it is not a central point, replacement methods do mesh well with the complex assessments that are frequently undertaken to retrospectively assign exposures in epidemiological studies.

The LSS served as a template for the simulation conditions. Because of the large sample in this study, the cases and person-years of risk are often grouped into categories and analyzed with a Poisson regression model. Hence a Poisson regression was used in the simulations. Typically these regression models incorporate effects for city, sex, age, and birth cohort (age in 1945) into the background disease rates and use a linear model in the radiation dose. For the purposes of the simulations, this was parameterized as $\alpha + \beta x$. The simulation used a lognormal model for measurement errors and skewed distributions for the true covariates, x .

Of course, the simulations differ from the LSS in several ways; two that are noteworthy being the sample size and grouping of data. The simulated sample sizes were much smaller due to the tremendous amount of computing that a simulation using a sample of 76,000 would require. Also the simulation did not group data; grouping poses few conceptual or practical problems with a replacement model and linear links but other methods, e.g., corresponding maximum likelihood models, are technically difficult to consider when the data are grouped.

3.2 A Measure of Relative Efficiency

Traditionally, the appraisal of a measurement error method has focused on its ability to remove or reduce the bias present in the naive estimates. This can be too crude a criterion to adequately distinguish among several good estimators and, for this reason, the emphasis in this work is on the relative efficiency of estimators. Most of the time the relative efficiency cannot be calculated and employed directly to rank methods because the efficiency depends upon unknown distributions. However, consideration of each method's relationship to the quasi-structural model can provide a substitute assessment of the relative efficiency.

Two generic techniques for approximating the quasi-structural score were introduced in Section 2.4. These techniques were respectively exemplified by the Nakamura method (Section 2.6) and replacement methods (Section 2.7). The Nakamura method will furnish consistent parameter estimates and usually eliminates the bias problem. However, it can be less efficient than a method that obtains consistent parameter estimates through more specific information about $\mathrm{F}_X^n(\cdot)$. This results because the Nakamura method's approximation to the quasi-structural score is usually less accurate. A replacement method offers this potential improvement in efficiency over the Nakamura estimator. The difficulty of a replacement method is that error in the specification of $\mathrm{G}(\cdot)$ will usually yield an estimator which is not consistent.

The best estimate that any measurement error method can hope to recover is the estimate that would be obtained if there were no measurement errors. In the class of generalized linear models under consideration, this is the maximum likelihood estimate, $\overset{*}{\hat{\varphi}}(\mathbf{x})$. Accordingly, a good standard of how well an arbitrary estimator, $\breve{\varphi}$, rectifies the problems that have been introduced by the measurement errors is the magnitude of the expected squared difference,

$$\mathrm{E}_X\left[\mathrm{E}_{YZ|X}\left[\breve{\varphi}-\overset{*}{\hat{\varphi}}(\mathbf{x})\right]^2\right]. \qquad 3.2\text{-}1$$

For our purposes, these expectations will be estimated through a simulation. In a functional setting, $\mathbf{x}$ is constant over each simulation trial and, in the structural

setting, $\mathbf{x}$ will be a sample from $F_X(\cdot)$ on each simulation trial. An important aspect of the calculation is blocking on the estimators from each data set, $(\mathbf{x},\mathbf{y},\mathbf{z})_t$. That is, on the t "th" trial, our interest is in the difference, $\left|\breve{\varphi}_t - \overset{*}{\hat{\varphi}}_t(\mathbf{x})\right|$, and the calculated expectation will be a function of

$$\sum_t \left[\breve{\varphi}_t - \overset{*}{\hat{\varphi}}_t(\mathbf{x})\right]^2 . \qquad 3.2\text{-}2$$

Since $\hat{\varphi}(F_X^n)$ and $\overset{*}{\hat{\varphi}}(\mathbf{x})$ are defined for both functional and structural models, the relative efficiency of any estimator of a component of φ, say $\breve{\beta}$, can be calculated in the simulations as

$$\text{eff}(\breve{\beta}) = \frac{\sum_t \left[\hat{\beta}_t(F_X^n) - \overset{*}{\hat{\beta}}_t(\mathbf{x})\right]^2}{\sum_t \left[\breve{\beta}_t - \overset{*}{\hat{\beta}}_t(\mathbf{x})\right]^2} . \qquad 3.2\text{-}3$$

This definition of relative efficiency relates the ability of an estimator to recover the "with no error" estimator, $\overset{*}{\hat{\varphi}}(\mathbf{x})$, to that of the quasi-structural maximum likelihood estimator, $\hat{\varphi}(F_X^n)$. This numerator simply provides a convenient scaling and there is no need to assume that $\text{eff}(\breve{\beta}) \le 1$. However, if the distribution of the true covariate values is an important determinant of efficiency then the quasi-structural maximum likelihood estimator, $\hat{\varphi}(F_X^n)$, should be very efficient. Since the estimating equation for many methods can be thought of as an approximation to this estimator's score, this definition relates relative efficiency to the degree of approximation implied by a method. Among replacement methods, this means that efficient estimation hinges on how well the empirical distribution of true covariate values is approximated. Since Nakamura methods often lead to estimators that are equivalent to replacement estimators a similar evaluation also pertains to them.

3.3 Simulation Study

Simulation experiments were conducted to clearly show through some examples that sizable gains in efficiency can accompany the selection of an estimation method and that attention to the distribution of the true covariate values is a key to achieving this increase in efficiency. The simulations emulate conditions that occur in the Life Span Study, but it was not the intention to exhaustively vary these conditions. Rather, a few cases are presented to demonstrate how the quasi-structural score equations can guide the selection of an estimation method in the context of an epidemiological cohort study.

3.3.1 Methods

A Poisson model with mean, $\mu = 0.5 + \beta x$, was used for $f_Y(\cdot;\mu)$; a lognormal distribution with coefficient of variation, ν, was used for $f_{Z|X}(\cdot)$; and a mixture of two lognormal distributions or a single lognormal distribution was used for $F_X(\cdot)$. Eight series of 2500 data sets were created; 500 data sets with samples of 100, 200, 400, 800, and 1600 observations. Each series was defined by the value for the slope, β, the coefficient of variation, ν, and the distribution for the unobserved covariates, $F_X(\cdot)$. That is, a triplet, $(\beta, \nu, F_X(\cdot))$, identifies each series. The specific values that were used for $(\beta, \nu, F_X(\cdot))$ in the eight series are listed in Table 3.

A triplet, $(\beta, \nu, F_X(\cdot))$, such as in the first series of Table 3 will be referenced in the results and their discussion as (0.1,0.3,L) where "L" refers to the lognormal distribution for $F_X(\cdot)$ and a triplet, such as the last series, will be referenced as (0.2,0.5,M) where "M" refers to the mixture distribution for $F_X(\cdot)$. These two distributions for $F_X(\cdot)$ are plotted in Figure 1. They have the same first two moments with the mixture distribution being more skewed than the lognormal distribution.

In two of the series, (0.1,0.3,M*) and (0.1,0.5,M*), the 500 data sets share a single vector, $\mathbf{x}$, which was a sample from the mixture distribution. These two series mimic the functional setting while the other six series mimic the structural setting by randomly sampling the distribution thereby using a different $\mathbf{x}$ in each data set.

Estimates were not calculated at $n = 100$ in the two series involving samples from the lognormal distribution. These were dropped because of the frequency with which negative values for $\mu = \alpha + \beta x$ occurred. This problem did not arise in the simulations involving the mixture model or simulations at the larger sample sizes.

Ten methods were used to estimate the intercept and slope for each data set. The general definition of these estimators is given in Chapter 2 and the details involved in adapting each of these methods to this Poisson - multiplicative measurement error model are given in Chapter 4. The slope estimators were:

1. the maximum likelihood estimate for the Poisson model without measurement error, which will be denoted as $\overset{*}{\beta}(\mathbf{x})$;
2. the maximum likelihood estimate for the quasi-structural model, $\hat{\beta}(\mathrm{F}_X^n)$;
3. the iteratively reweighted least squares estimate for the quasi-structural model, $\tilde{\beta}(\mathrm{F}_X^n)$;
4. the replacement estimator for the quasi-structural model, $\overset{*}{\beta}(\xi_{\mathrm{F}_X^n})$, i.e., based upon $\xi_{\mathrm{F}_X^n}$;
5. the iteratively reweighted least squares estimate for the structural model, $\tilde{\beta}(\mathrm{F}_X)$,
6. a Whittemore estimate, $\tilde{\beta}(\xi_{\mathrm{G}})$, adapted to lognormal distributions for $\mathrm{f}_{Z|X}(\cdot)$ and $\mathrm{F}_X(\cdot)$;
7. Schafer's quadratic estimator for multiplicative error, $\overset{*}{\beta}(\xi_{\mathrm{q}})$;
8. an approximate Nakamura estimator, $\widehat{\beta}$,
9. a version of a correction for attenuation, $B\overset{*}{\beta}(\mathbf{z})$;
10. and the naive estimate, $\overset{*}{\beta}(\mathbf{z})$.

The simulations estimated the efficiency of the last nine estimators where the relative efficiency of a given estimator, $\breve{\beta}$, was calculated as Equation 3.2-3,

$$\mathrm{eff}\left(\breve{\beta}\right) = \frac{\sum_t \left[\hat{\beta}_t\left(\mathrm{F}_X^n\right) - \overset{*}{\hat{\beta}}_t(\mathbf{x})\right]^2}{\sum_t \left[\breve{\beta}_t - \overset{*}{\hat{\beta}}_t(\mathbf{x})\right]^2} \,.$$

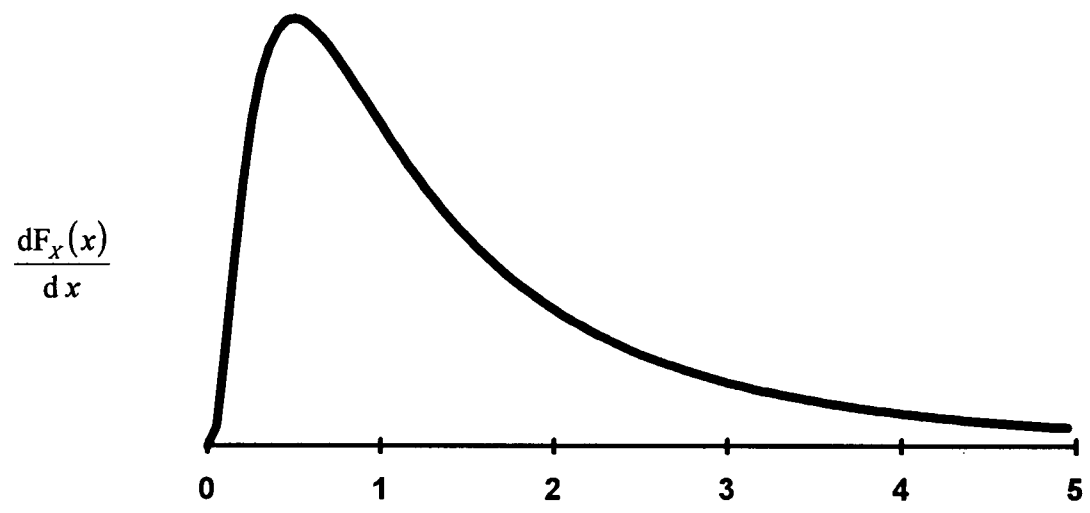
The summation extends over 500 data sets from a given series and sample size. By definition the relative efficiency of the maximum likelihood estimate for the quasi-structural model is 1, i.e., $\mathrm{eff}\left(\hat{\beta}\left(\mathrm{F}_X^n\right)\right)$, and the other estimators are compared to it. A motivation for this definition of relative efficiency is contained in Section 3.2.

Table 2. Parameter Values Used in Each Series of Simulations

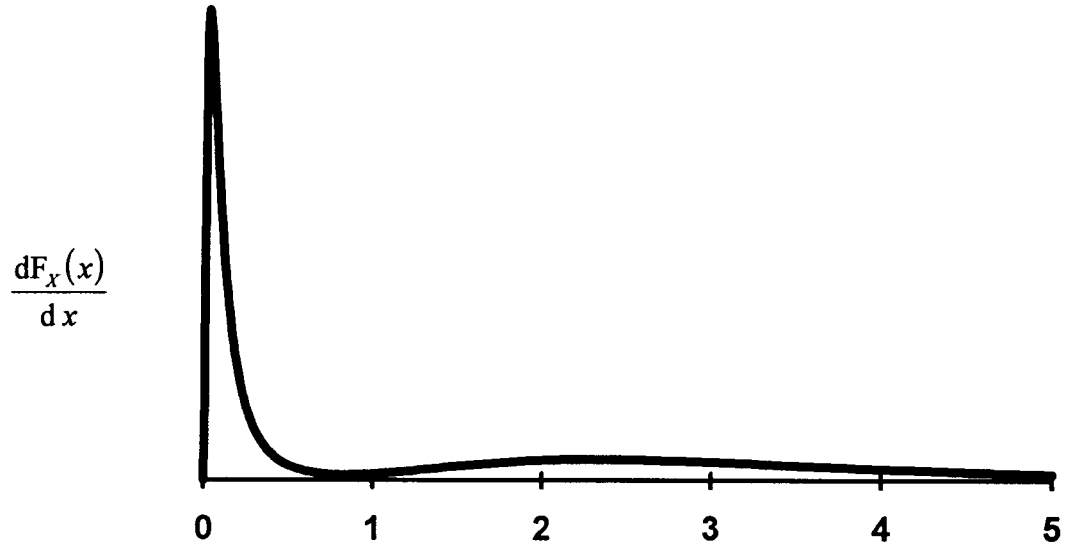
			$F_X(\cdot)$			
Series	β	ν	μ_1	σ_1	μ_2	σ_2
(0.1,0.3,L)	0.1	0.3	0.151	0.924		
(0.1,0.5,L)	0.1	0.5	0.151	0.924		
(0.1,0.3,M)	0.1	0.3	0.1	1.0	3.0	0.5
(0.1,0.5,M)	0.1	0.5	0.1	1.0	3.0	0.5
(0.1,0.3,M*)	0.1	0.3	0.1	1.0	3.0	0.5
(0.1,0.5,M*)	0.1	0.5	0.1	1.0	3.0	0.5
(0.2,0.3,M)	0.2	0.3	0.1	1.0	3.0	0.5
(0.2,0.5,M)	0.2	0.5	0.1	1.0	3.0	0.5

* All data sets in these series use a single sample, $\mathbf{x}$, from $F_X(\cdot)$.

Figure 1. Simulated Distributions for *x*



Lognormal



Mixture

3.3.2 Results and Discussion

The relative efficiencies of the nine estimators for the slope, $\mathrm{eff}\left(\tilde{\beta}\right)$, are presented as Tables 3 - 10. The intercept estimators essentially duplicate the relative efficiency of the slope estimators so they are not reported.

There are two clear trends across these tables. The relative efficiency of each estimator increases as either the coefficient of variation decreases or the slope decreases. Such trends are predictable because decreasing either the slope or the coefficient of variation decreases the measurement error and its effects. As $\mathbf{z} \to \mathbf{x}$ all of these estimators converge to $\overset{*}{\beta}(\mathbf{x})$. Incorporating this fact into the respective approximations to the quasi-structural score implies that decreasing the measurement error will increase the correlation among these estimators and, consequently, their relative efficiencies.

The approximation to the quasi-structural score suggests some other patterns that are also apparent in the tables. Estimators that use the empirical distribution are the more efficient and the maximum likelihood estimator for the quasi-structural model, $\hat{\beta}\left(\mathrm{F}_X^n\right)$, is, as expected, the most efficient. Under the simulated conditions, the iteratively reweighted least squares estimator, $\tilde{\beta}\left(\mathrm{F}_X^n\right)$, and the replacement estimator, $\overset{*}{\beta}\left(\xi_{\mathrm{F}_X^n}\right)$, are nearly as efficient as the maximum likelihood estimator. The high relative efficiency of these two estimators follows from their approximation to the quasi-structural score. Whenever y_i contains little information about x_i, the case in these simulations, then $\mathrm{E}_{X|YZ}(\cdot) \approx \mathrm{E}_{X|Z}(\cdot)$ and the estimating equations of these two estimators almost duplicate the quasi-structural score. Although the approximation is very close, the computation involved in solving these estimating equations is very different. This observation is at the heart of the recommendation favoring the use of replacement methods in cohort studies.

In small samples, the maximum likelihood estimator for the quasi-structural model is expected to average closer to $\overset{*}{\beta}(\mathbf{x})$ than the structural maximum likelihood estimator. The structural maximum likelihood estimator was not calculated but this prediction is indirectly corroborated by the respective iteratively reweighted least

squares estimators. Since $\tilde{\beta}(F_X^n)$ is nearly as efficient as $\hat{\beta}(F_X^n)$, a comparison of the iteratively reweighted least squares estimators for the quasi-structural model, $\tilde{\beta}(F_X^n)$, and the structural model, $\tilde{\beta}(F_X)$, will replicate a comparison of the respective maximum likelihood estimators. Overall, the structural model iteratively reweighted least squares estimator, $\tilde{\beta}(F_X)$, is relatively efficient. However this estimator is always less efficient than the quasi-structural estimator, $\tilde{\beta}(F_X^n)$, and in some cases it is significantly less efficient.

Since the difference between $F_X^n(\cdot)$ and $F_X(\cdot)$ will diminish as samples become large, estimators based upon these distributions become equivalent. At least in the structural model simulations (Tables 3 - 6, 9, 10), there is a hint of this prediction with increasing sample size although sampling variation in these simulations apparently blurs the shallow trends.

As a practical matter, the best surrogate for $F_X^n(\cdot)$ would appear to be $F_X(\cdot)$. Thus, the efficiency of $\hat{\beta}(F_X)$ is conceptually a bound on the efficiency of methods based upon picking a distribution, $G(\cdot)$. The methods that can be used in practice, i.e., methods 6-10 as listed in Section 3.3.1, are typically less efficient than this method assuming that its efficiency is evidenced by $\tilde{\beta}(F_X)$. In "small measurement error" cases, i.e., essentially when $\nu = 0.3$, the better among the practicable estimators approach this bound. Occasionally, the efficiency of $\tilde{\beta}(F_X)$ is exceeded although this is believed to reflect sampling variability as well as a small difference from the efficiency of $\hat{\beta}(F_X)$.

Two general patterns attend increases in sample size. Several estimators exhibit little change in efficiency over the range of sample sizes. The quasi-structural replacement estimator or the quasi-structural iteratively reweighted least squares estimator give the best examples of this pattern although the Schafer estimator and the Nakamura estimator typically display this pattern. Other estimators, at least in some of the tables, exhibit a stair-step decline in efficiency with increasing sample size. This pattern is expected whenever an estimator is not consistent for β. For an inconsistent estimator, the decline in efficiency reflects the increasing dominance of a bias component in the mean squared error. The naive estimator exemplifies this

phenomenon. Although less clear cut, the stair-step pattern usually occurs with the correction for attenuation estimator. This pattern regularly occurs with the modified Whittemore estimator when $F_X(\cdot)$ is the mixture of lognormal distributions (Tables 5 - 10).

The naive use of $\mathbf{z}$ is usually not considered to be a measurement error method. However, the estimating equation for this estimator can be viewed as an approximation to the quasi-structural score. As such, it is simply a poorer approximation than used by the other methods. Compared to the correction for attenuation estimator or the Nakamura estimator, the naive estimator ignores all but the zero order term in a Taylor's expansion of the quasi-structural score and, compared to replacement estimators, it approximates the expected value of x given z with z . Either approximation should be improved by incorporating other information and there is no surprise in the naive method's inferior performance. What is clear in this viewpoint of the naive estimator as a replacement estimator is that there are potentially worse replacement estimators. They could conceptually be constructed from distributions, $G(\cdot)$, that grossly misrepresent $F_X(\cdot)$. This, in fact, is the flaw with the Whittemore estimator when used in the mixture model for $F_X(\cdot)$, Tables 5 - 10.

From the standpoint of bringing into harmony both the requirements of cohort studies and the requirements of measurement error models, replacement methods are very attractive. These simulations were presented to illustrate that simply selecting a convenient replacement method with only a superficial concern for how well it approximates $F_X^n(\cdot)$ can give nearly as poor an estimate as the naive estimate. The method of Whittemore as modified to lognormal distributions was selected to be the straw man. In the simulations reported as Tables 3 and 4, the modified Whittemore method should be very efficient since it correctly assumes that $f_{Z|X}(\cdot)$ and $F_X(\cdot)$ are lognormal distributions. In the remaining simulations, the modified Whittemore method would appear to be a reasonable method, at least superficially; it adjusts for the known measurement error distribution, $f_{Z|X}(\cdot)$, and it can accommodate a skewed distribution for $F_X(\cdot)$. However, as demonstrated by the simulations, this estimator is no better than the naive estimator.

The Whittemore algorithm assumes that $F_X(\cdot)$ is a lognormal distribution. The idea behind using a mixture to thwart this algorithm is depicted in Figure 2. As the method is defined in Equation 2.7.1-6, the $\ln \xi_G(z)$ is linear in z and Figure 2 sketches this relationship for the mixture distribution and the constituent lognormal distributions. The "S" shaped curve is the true relationship that results from the mixture distribution. Presumably the Whittemore algorithm must compromise by "running a diagonal" between the two dashed lines in an attempt to model the true shape. The resultant estimates, $\xi_G(z)$, are in error almost everywhere. The simulation definitely confirms poor performance.

Another motive for the simulations was to clearly illustrate that sizable gains in efficiency can accompany the selection of an estimation method. The modified Whittemore method, in particular, shows this. It is efficient under the lognormal distribution for $F_X(\cdot)$ and is little better than no correction under the mixture distribution. In these latter simulations, alternate methods can be several times more efficient. Methods like Schafer's or Nakamura's are more stable throughout these simulations but this should not be over interpreted since both rely on assumptions that could also be manipulated to their disadvantage. In particular, replacement estimators are easy to sabotage in a simulation. The quadratic Schafer estimator, for example, approximates the expected value of x given z with a quadratic function in z . Manipulating $F_X(\cdot)$ so that the expected value of x given z is not a quadratic function in z is not too difficult. In these simulations, the converse manipulation is evidenced in Tables 7 and 8 where a vector, $\mathbf{x}$, was found through trial and error such that $\xi_{F_X^n}(z)$ was well approximated by a quadratic function in z . Especially in the simulation reported as Table 7, this manipulation boosted the efficiency of the Schafer method over that of the Nakamura method, a method that otherwise had a similar efficiency.

Unlike the replacement methods, the Nakamura method is somewhat robust to manipulation of $F_X^n(\cdot)$. A cost of robustness is that a judiciously chosen replacement method will be more efficient. This is seen in Tables 3 and 4 where the Whittemore method is more efficient and in Tables 7 and 8 where the Schafer method is more efficient. While the differences in efficiency illustrated in these tables are not large, experience with other simulations involving normal regression models has demonstrated more substantial differences. A limitation operating in these examples

is that the best replacement method, $\hat{\beta}^{*}\left(\xi_{F_X^n}\right)$, is not substantially more efficient. Under conditions where there is more room for improvement, such as simulated in Table 8, a well chosen replacement estimator could easily be twice as efficient as the Nakamura estimator. The Schafer method and the Whittemore method do not perform well in Table 8 because these estimators are biased estimators of β in the circumstances simulated. Other practical concerns in adopting a Nakamura method, include the requirement of special programs to implement the method and the lack of transparency especially to the selection of a regression model.

Since the Nakamura method is a refinement of the correction for attenuation method, the correction for attenuation is expected to be less efficient. This is true for these simulations and one would recommend against this method whenever it differs from the Nakamura method. This method is usually proposed when the measurement errors and the regression model are normally distributed. Under those circumstances these two methods are equivalent.

Figure 2. Expected Value of x given *z* for the Mixture Distribution and its Component Lognormal Distributions.

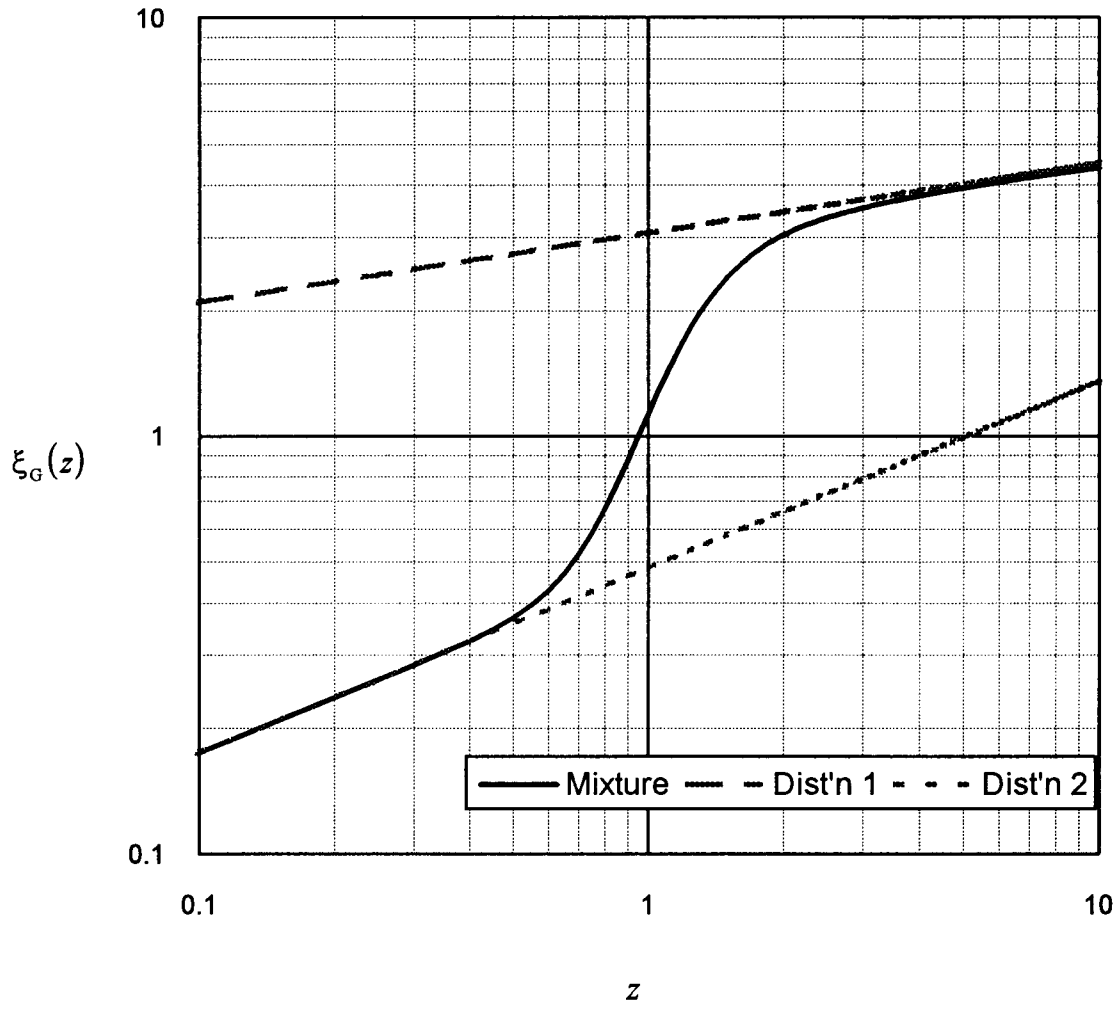


Table 3. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.3,L).

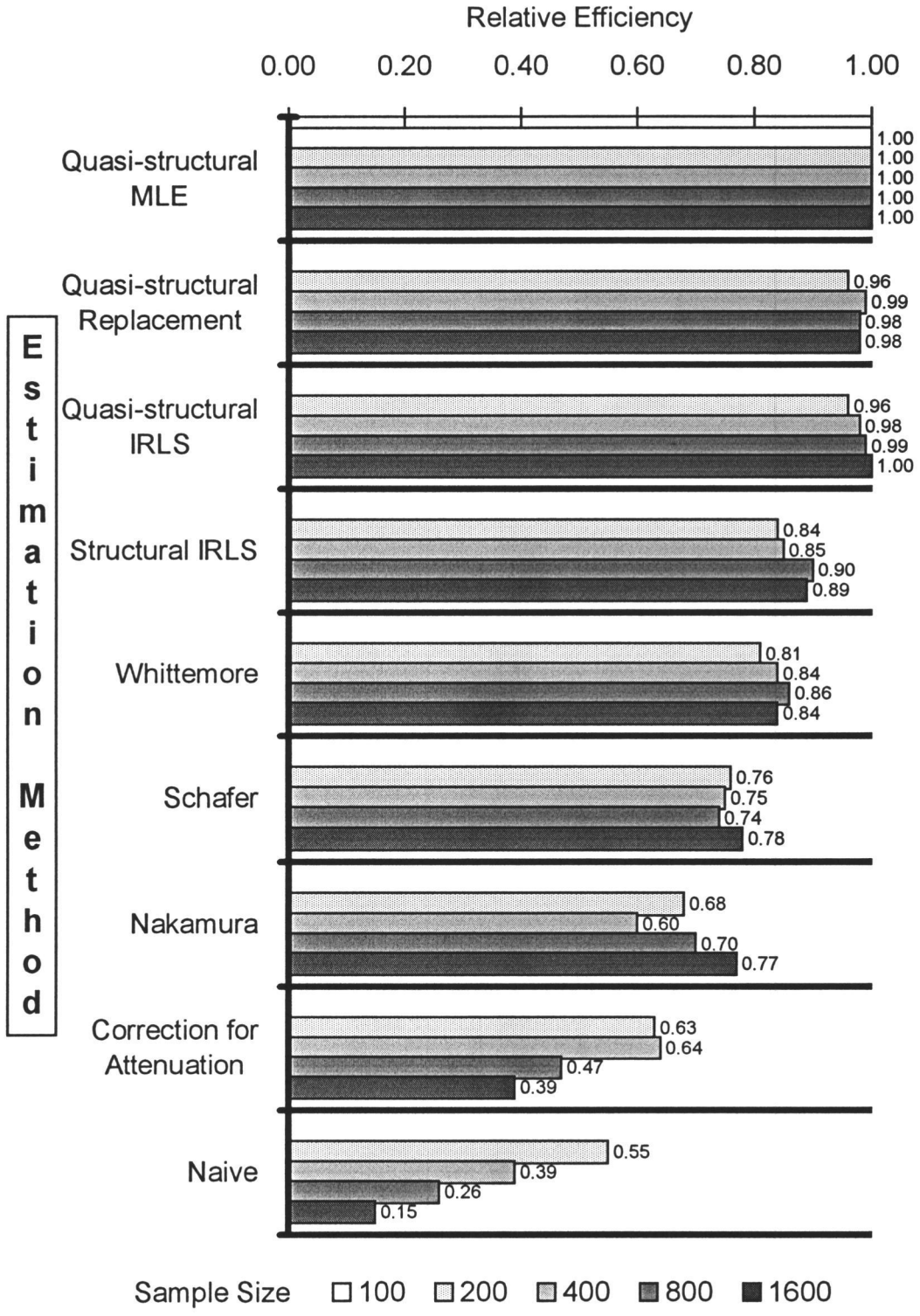


Table 4. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.5,L).

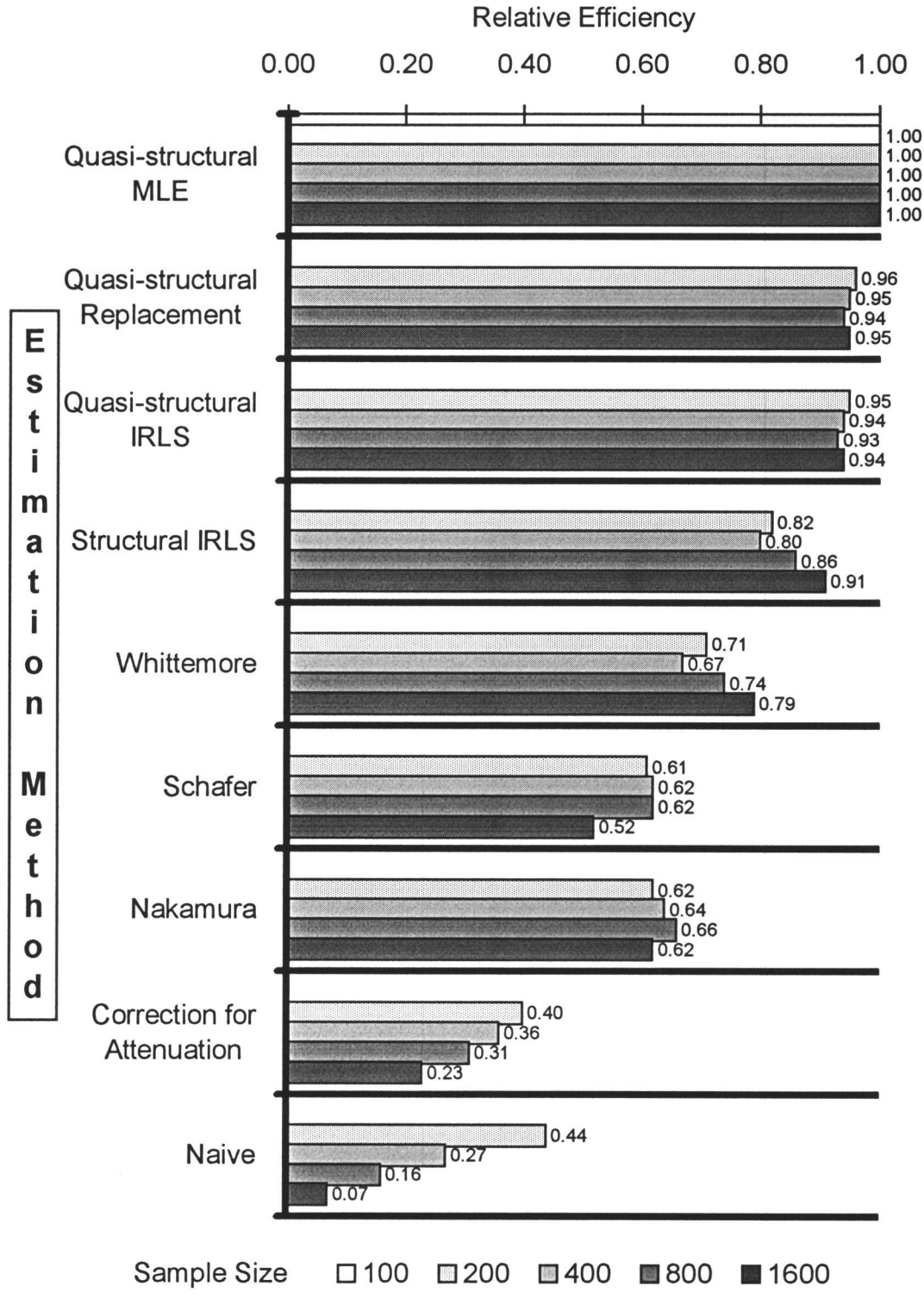


Table 5. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.3,M).

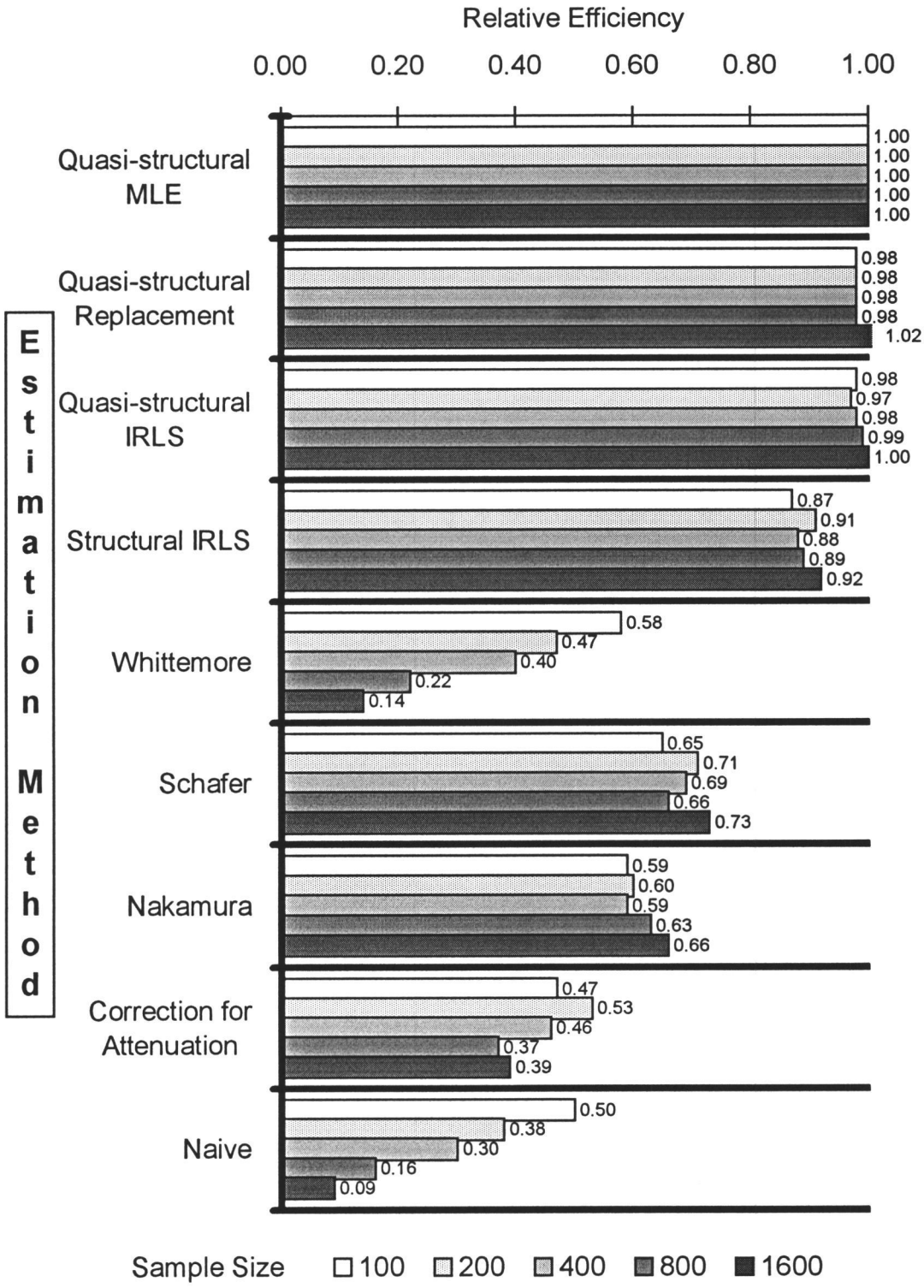


Table 6. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.5,M).

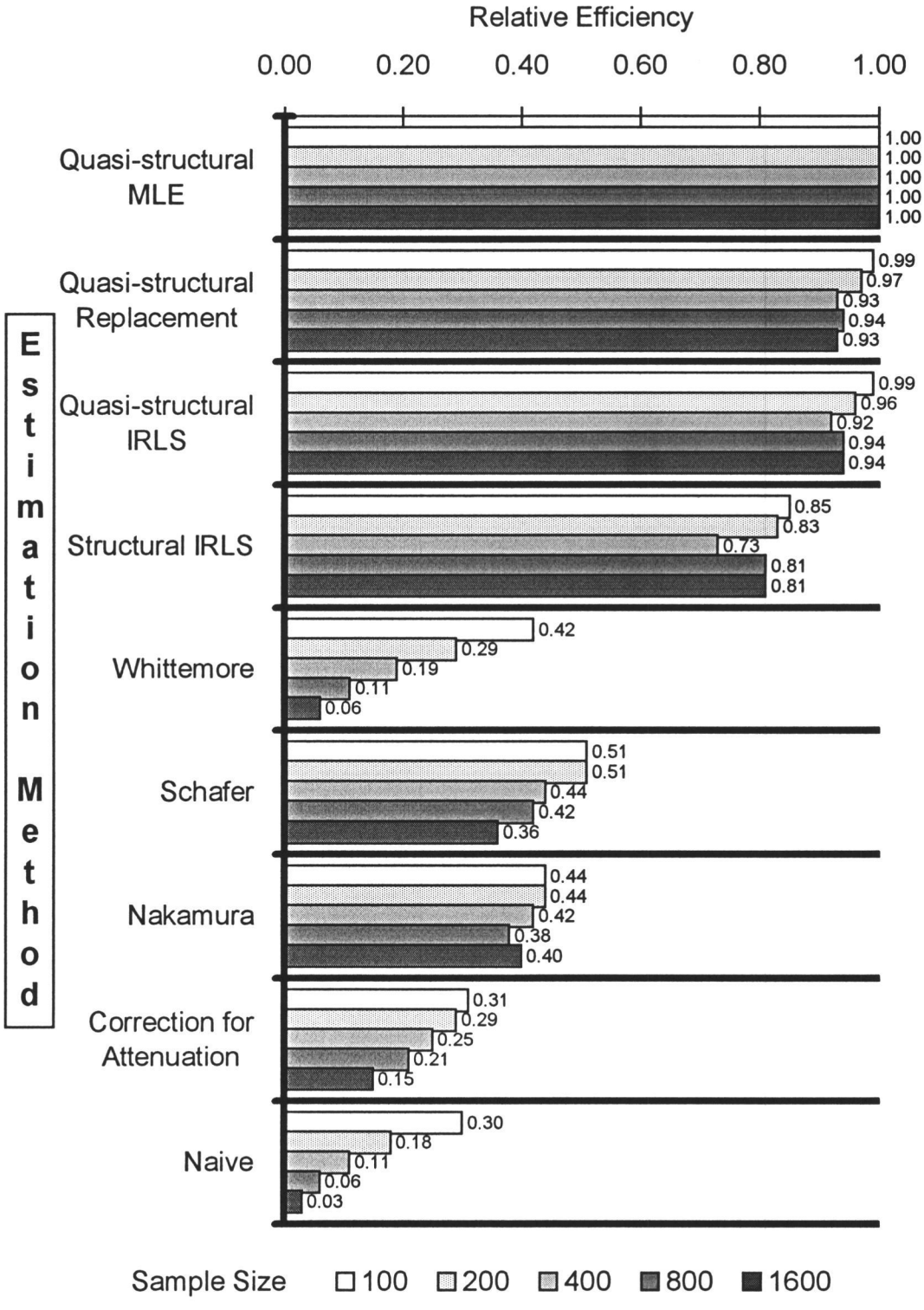


Table 7. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.3,M*).

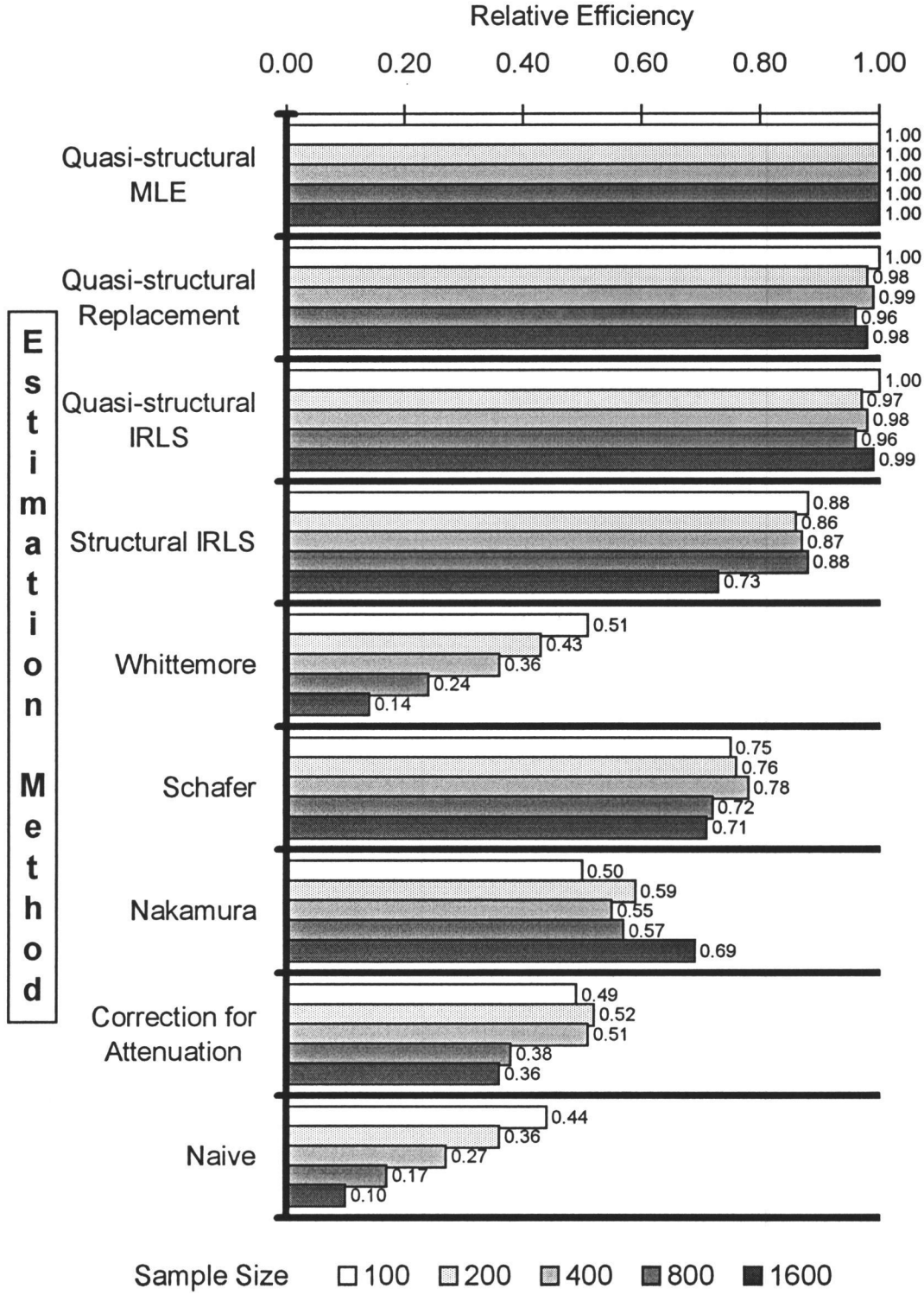


Table 8. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.1,0.5,M*).

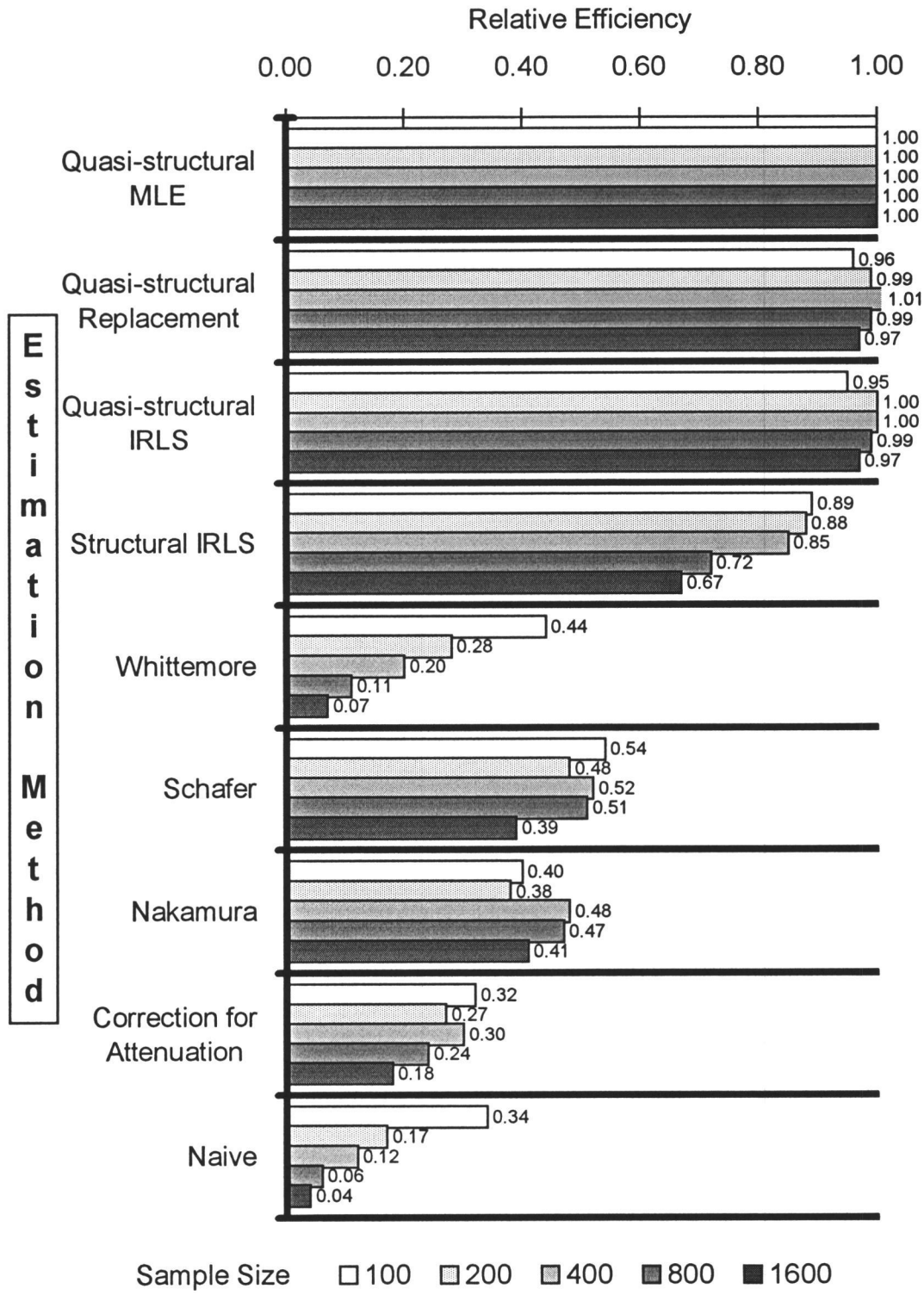


Table 9. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.2,0.3,M).

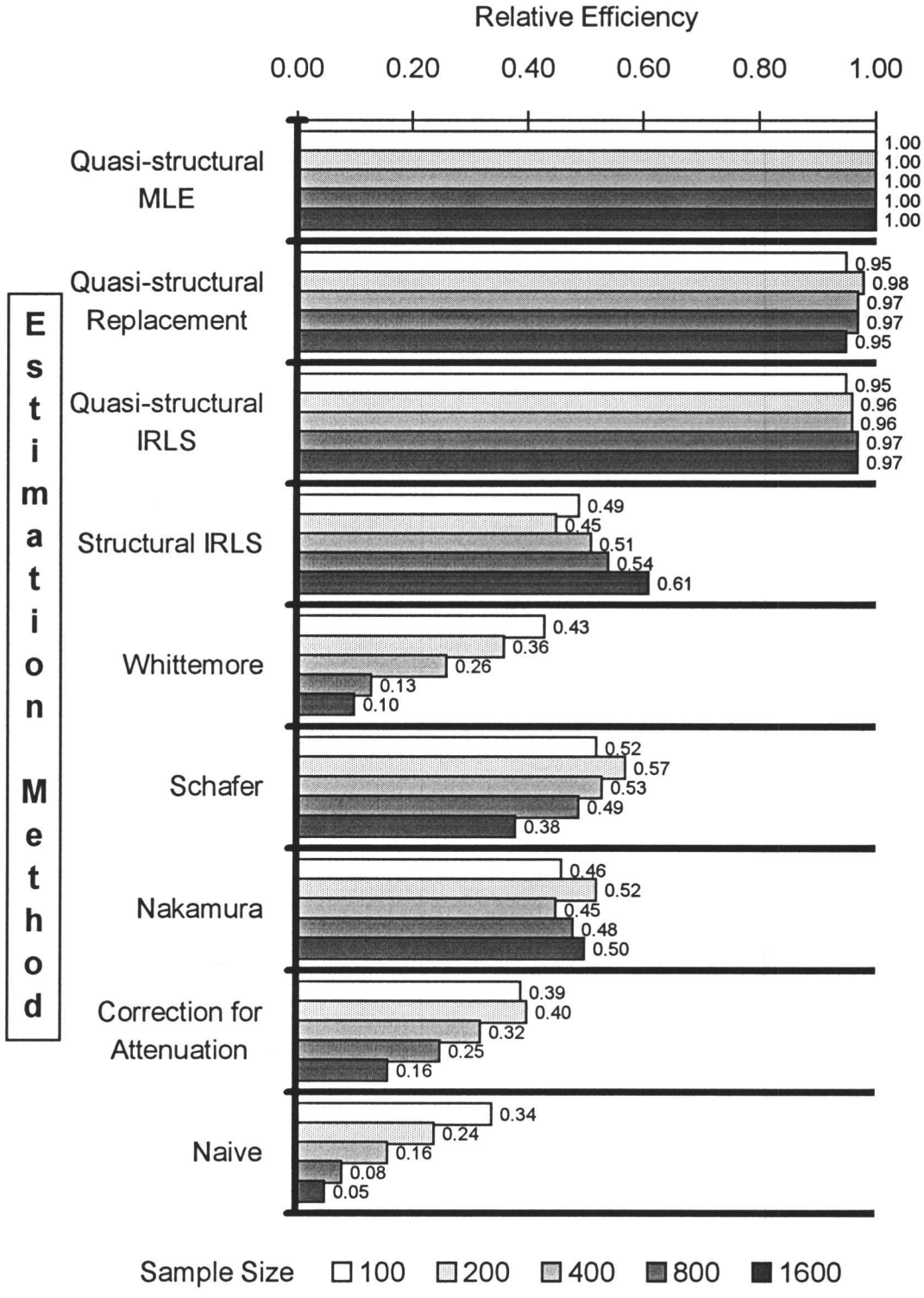
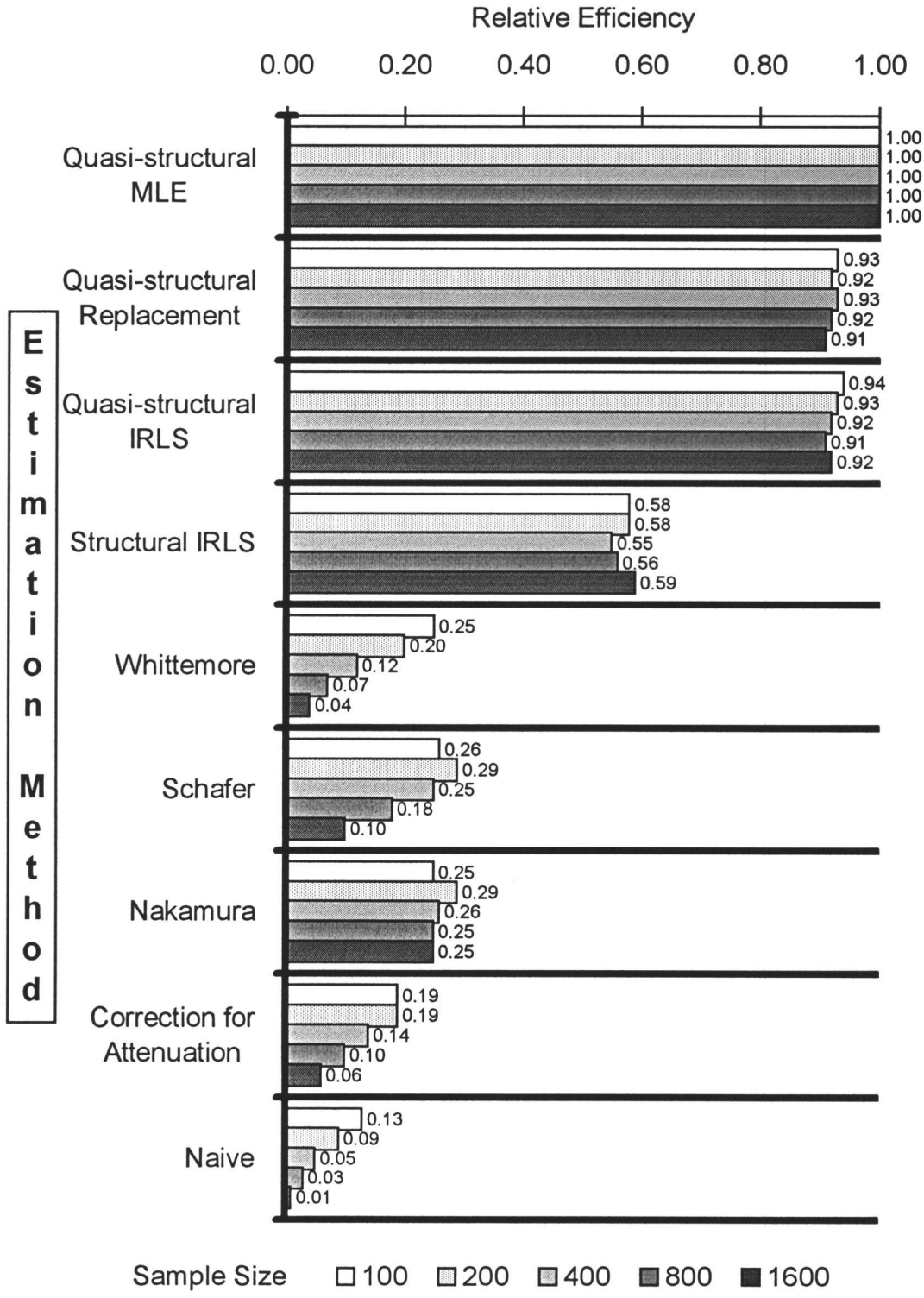


Table 10. Comparison of the Relative Efficiencies of Several Estimators for the Slope under the Model (0.2,0.5,M).



3.4 Conclusions

The principal premise of this dissertation is the importance of explicitly and carefully considering the distribution of the true covariate values in regression models with covariate measurement error. In the functional setting, the empirical distribution of the true covariate values can guide this process through the quasi-structural model. That is, even though the quasi-structural maximum likelihood estimator is not directly applicable, it provides a goal that can be approximated. In particular, traditional estimators for the functional setting can be thought of as solutions to estimating equations that approximate the quasi-structural score. Since some methods make better approximations they are more efficient at correcting the problems introduced by measurement errors.

In this work, the relative efficiency of a method has a somewhat unconventional definition. Our interest in covariate measurement error problems is more focused upon how effectively a method recovers the estimate that would be obtained if $\mathbf{x}$ were observed, and takes for granted that the overall efficiency of a method partitions into two components; the efficiency with which the method recovers the estimate given $\mathbf{x}$ and the efficiency of the underlying method when $\mathbf{x}$ is known. Thus the relative efficiency of a method has been defined in a way that makes it virtually synonymous with the closeness of the approximation to the quasi-structural score implied by its estimating equation. The quasi-structural maximum likelihood estimator is presumably closer to the estimate that would be obtained if $\mathbf{x}$ were observed than any estimate based on less specific knowledge of $\mathbf{x}$.

Given this perspective, good estimators, in the sense that their estimating equation well approximates the quasi-structural score, can be obtained by finding a distribution, $G(\cdot)$, to replace the unobserved empirical distribution, $F_X^n(\cdot)$. In this endeavor, the size of the sample becomes important since the main insight into the empirical distribution of the unobserved covariates comes from the empirical distribution of the observed covariates, $F_Z^n(\cdot)$ given $f_{Z|X}(\cdot)$. In small samples, $F_Z^n(\cdot)$ is compatible with a broad class of surrogate distributions and consequently the specific choice is not so critical. However, as the sample is increased the class of distributions that are compatible with $F_Z^n(\cdot)$ given $f_{Z|X}(\cdot)$ becomes more distinct and the consequences of a particular choice become more substantial. This phenomenon was clearly demonstrated in the simulation, Tables 6 - 11, where the modified

Whittemore method has acceptable efficiency relative to the Schafer or Nakamura methods at a sample size of 100 and terrible efficiency at a sample size of 1600.

An alternate strategy is to adopt a procedure that can consistently estimate the regression without specifying $F_X^n(\cdot)$ or a surrogate, e.g., a correction for attenuation method or a Nakamura (1990) method. Generally such a method would have good relative efficiency in small samples since it is fairly unbiased and the class of distributions compatible with $F_Z^n(\cdot)$ given $f_{Z|X}(\cdot)$ is large, presumably encompassing any distribution implied by the method. In large samples, these estimates will be nearly unbiased and typically they will have better efficiency than structural methods that are predicated on an inappropriate $G(\cdot)$. However, their approximation to the quasi-structural score is crude, fundamentally approximating $F_X^n(\cdot)$ through the moments, $E_Z(z^k)$, given the assumed distribution, $f_{Z|X}(\cdot)$. These estimators are competitive when $F_X^n(\cdot)$ can be effectively summarized by a few moments as in the Gaussian examples of Section 2.5. But, they lose out, often dramatically, whenever the class of distributions compatible with $F_Z^n(\cdot)$ given $f_{Z|X}(\cdot)$ necessitates other moments, e.g., distributions that are highly skewed or have more than one mode.

As a rule, in epidemiological studies where sample sizes are very large and the class of distributions compatible with $F_Z^n(\cdot)$ given $f_{Z|X}(\cdot)$ are very non-Gaussian, the more efficient among potential estimation methods would be found by specifying a distribution for the true covariates. This dissertation does not address how this might be done but it does suggest that this problem can be reduced to estimating the expected value of x given z, $\xi_{F_X^n}(z)$. Whether this actually simplifies finding a suitable $G(\cdot)$ or complicates the process is unclear but it does place this aspect of the measurement error problem firmly within the framework of empirical Bayes estimation (Maritz and Lwin, 1989), and potentially opens a rich set of possibilities for incorporating a replacement for $F_X^n(\cdot)$. The quadratic Schafer estimator (1990) is one example of these possibilities. Since $F_Z^n(\cdot)$ is observed, a kernel density estimate for $f_Z(\cdot)$ can be constructed, and since $f_{Z|X}(\cdot)$ is assumed, $G(\cdot)$ could conceivably be produced as the deconvolution of this density. Recent work by Stefanski and Carroll (1990, 1991), for example, seems to be philosophically oriented in this direction. As a practical matter, such procedures may be unnecessarily sophisticated. In

applications, the specification of $f_{Z|X}(\cdot)$ will involve some guesswork, and it might be as valid to simply construct $G(\cdot)$ as a mixture of a few convenient distributions chosen to capture the salient properties of $F_X^n(\cdot)$. This is essentially the philosophy employed in the analyses of cancer incidence among the A-bomb survivors (Pierce *et al.*, 1992).

Under the prevailing conditions in cohort studies, the replacement estimator, $\hat{\hat{\varphi}}(\xi_G)$, will be nearly as efficient as the corresponding structural maximum likelihood estimator, $\hat{\varphi}(G)$. Assuming that $G(\cdot)$ is a satisfactory replacement for $F_X^n(\cdot)$, this observation forms the basis for recommending replacement or iteratively reweighted least squares approaches to correct for covariate measurement error in these studies. A small forfeiture of efficiency relative to the structural maximum likelihood estimator is outweighed by the relative versatility of this procedure in exploratory analyses as well as its ease of implementation. Replacement methods generally are more efficient than methods that avoid specifying a distribution for the true covariates, such as the method of Nakamura. Even in circumstances were a Nakamura estimator is a competitive alternative, the ease of use and versatility of replacement methods give them a significant edge.

Chapter 4. Details of the Estimators and the Methods Used in the Simulations

This chapter documents the more technical aspects of the simulations that were reported in Chapter 3. A detailed description is given of the distributions that were sampled, the computer methods that were employed, and the statistical methods that generated the estimates. Such detail was avoided in earlier sections to stress the principles underlying these methods. As a result some of the estimators that were employed in the simulations may need more explicit definition.

The simulations roughly reflect the Life Span Study in the sense that they use a Poisson model with an identity link as the model for $f_Y(\cdot;\mu)$, i.e. $\mu = \alpha + \beta x$; a lognormal distribution with constant coefficient of variation, ν, for $f_{Z|X}(\cdot)$; and a distribution for $F_X(\cdot)$ which is skewed to favor low exposures. Of course, the actual situation in the Life Span Study is more complex but these choices for the distributions focus the simulation experiments on modeling issues related to the distribution of the unobserved covariate.

Ten estimators were calculated for each of the simulated data sets. Five of these estimators are of theoretical interest inasmuch as they use information that would usually be unavailable. These "theoretical" estimators are:

1. the maximum likelihood estimator without measurement errors (i.e., using $\mathbf{x}$),
2. the quasi-structural maximum likelihood estimator (i.e., using $F_X^n(\cdot)$),
3. the quasi-structural iteratively reweighted least squares estimator,
4. the quasi-structural replacement estimator
5. and the structural iteratively reweighted least squares estimator (i.e., using $F_X(\cdot)$).

The remaining five estimators can be applied to real data. These estimators are:

6. the naive estimator,
7. a correction for attenuation estimator,
8. an approximate Nakamura estimator,
9. a modified Whittemore estimator
10. and the Schafer quadratic estimator.

4.1. Simulated Distributions

Two specific distributions for $F_X(\cdot)$ were used in the simulations. The first distribution was a mixture of two lognormal probability densities. Half of the elements in the unobserved vector, $\mathbf{x}$, came from the lognormal density with parameters, $\mu = 0.1$ and $\sigma = 1.0$, and the remainder came from the lognormal density with parameters, $\mu = 3.0$ and $\sigma = 0.5$. The lognormal density was parameterized as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} \exp\left\{ -\frac{(\ln(x) - \mu)^2}{2\sigma^2} \right\}. \qquad 4.1\text{-}1$$

This mixture distribution is displayed in Figure 1.

The second distribution was simply a lognormal distribution with the parameters, $\mu = 0.151$ and $\sigma = 0.924$. This distribution has the same mean and variance as the mixture. This density is also displayed in Figure 1.

It was convenient to take $F_X(\cdot)$ to be a mixture of lognormal distributions because this mixture eased calculation while permitting sufficient flexibility in the form of the distribution to have the heavy skewing illustrated in Figure 1. With respect to calculation, the quantities, $\xi_G(z)$ and $V_G(z)$, which are required for the iteratively reweighted least squares estimates in the structural model, have manageable closed form solutions (Appendix 1, Section A1-4). Pseudo-random samples were generated from this distribution by calling the RNDN function in the Gauss programming language.

The simulations used a multiplicative measurement error model with a constant coefficient of variation, ν. This specification requires that the mean and variance satisfy

$$E_{Z|X=x}(z) = x \quad \text{and} \quad V_{Z|X=x}(z) = (\nu x)^2 . \qquad 4.1\text{-}2$$

These requirements were implemented through a lognormal distribution for $f_{Z|X}(\cdot)$. Equation 4.1-2 requires that $\mu = \ln(x) - \ln\left(\sqrt{\nu^2 + 1}\right)$ and $\sigma^2 = \ln(\nu^2 + 1)$.

Two values for the coefficient of variation were simulated, $\nu = 0.3$ and $\nu = 0.5$. A coefficient of variation close to 0.3 coincides with that of the Life Span Study. Initial simulation results recorded moderate differences among the efficiencies of the estimators with $\nu = 0.3$ so larger errors were also examined.

Given a vector, $\mathbf{x}$, the response, $\mathbf{y}$, was generated as a pseudo-random sample from the Poisson distribution with mean, $\mu_i = 0.5 + \beta x_i$ .The slope was either $\beta = 0.1$ or $\beta = 0.2$ in these simulations.

The $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ data sets with the specified distributional properties were created in a Gauss program; XYZ.GSS, Appendix 2, Section A2-1.1. Lognormal samples were generated from standard normal pseudo-random deviates generated with the Gauss function, RNDN. That is,

$$\mathbf{z} = \exp(\mu + \mathrm{RNDN}(n,1) * \sigma) \qquad 4.1\text{-}3$$

Pseudo-random Poisson values were obtained by the PIF algorithm that is presented in Fishman (1976). The computer programs that implement this algorithm are PIF.GSS (Section A2-1.2) and QTABLE.FOR (Section A2-1.3).

4.2 Estimators for the Quasi-structural Model

In this section three estimators are defined for the quasi-structural model. They are the maximum likelihood estimator, Section 4.2.1; the iteratively reweighted least squares estimator, Section 4.2.2; and the replacement estimator, Section 4.2.3. These estimators follow directly by taking $f_Y(\cdot;\mu)$ to be a Poisson distribution with mean, $\mu = \alpha + \beta x$, and $f_{Z|X}(\cdot)$ to be a lognormal distribution with $E_{Z|X=x}(z) = x$ and constant coefficient of variation, ν . in the definitions given in Section 2.2.

4.2.1 Maximum Likelihood Estimates

The likelihood for this quasi-structural model, $L(\alpha,\beta;F_X^n,\mathbf{y},\mathbf{z})$, is defined by Equation 2.2.1-1. It follows directly by substituting the explicit component distributions into the definition that this likelihood satisfies

$$\ln L(\alpha,\beta;F_X^n,\mathbf{y},\mathbf{z}) = \textit{constant} + \sum_{i=1}^{n} \ln\left(\frac{w_{i\cdot}}{c_{i\cdot}}\right) \qquad 4.2.1\text{-}1$$

where

$$c_{ij} = \exp\left\{-\frac{\left(\ln(z_i)-\mu_j\right)^2}{2\sigma^2}\right\} \qquad 4.2.1\text{-}2$$

and

$$w_{ij} = c_{ij}\, e^{-(\alpha+\beta x_j)}\left(\alpha+\beta x_j\right)^{y_i} . \qquad 4.2.1\text{-}3$$

For the constant coefficient of variation error model, Equations 4.1-2,

$\mu_j = \ln(x_j) - \ln\left(\sqrt{\nu^2+1}\right)$ and $\sigma^2 = \ln(\nu^2+1)$.

The score equation for this model can be written as

$$\mathbf{U}(\alpha,\beta;F_X^n,\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} \overset{*}{\mathbf{U}}_i(\alpha,\beta;x_j,y_i)\, w_{ij} \Big/ w_{i\cdot}\right) \qquad 4.2.1\text{-}4$$

where

$$\overset{*}{\mathbf{U}}_i(\alpha,\beta;x_j,y_i) = \left(\frac{y_i}{\alpha+\beta x_j} - 1\right)\begin{bmatrix}1\\ x_j\end{bmatrix} . \qquad 4.2.1\text{-}5$$

The observed information matrix for this model is

4.2.1-6

$$\mathbf{i}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) = \begin{bmatrix} \mathrm{i}_{\alpha\alpha}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) & \mathrm{i}_{\alpha\beta}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) \\ \mathrm{i}_{\alpha\beta}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) & \mathrm{i}_{\beta\beta}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) \end{bmatrix}$$

where the components are defined by Equation 2.2.1-7. This gives the equations,

$$\mathrm{i}_{\alpha\alpha}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{i}}_{\alpha\alpha i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] - \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\alpha i}(\alpha,\beta;x_j,y_i)^2\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] + \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\alpha i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right]^2 \qquad 4.2.1\text{-}7$$

$$\mathrm{i}_{\alpha\beta}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{i}}_{\alpha\beta i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] - \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\alpha i}(\alpha,\beta;x_j,y_i)\,\overset{*}{\mathrm{U}}_{\beta i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] + \sum_{i=1}^{n}\left[\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\alpha i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right]\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\beta i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right]\right] \qquad 4.2.1\text{-}8$$

and

$$\mathbf{i}_{\beta\beta}(\alpha,\beta;\mathrm{F}_X^n,\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{i}}_{\beta\beta i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] - \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\beta i}(\alpha,\beta;x_j,y_i)^2\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right] + \sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n}\overset{*}{\mathrm{U}}_{\beta i}(\alpha,\beta;x_j,y_i)\,w_{ij}}{\sum_{j=1}^{n}w_{ij}}\right]^2 \qquad 4.2.1\text{-}9$$

where

$$\overset{*}{\mathbf{i}}_i(\alpha,\beta;x_j,y_i) = \frac{y_i}{(\alpha+\beta x_j)^2}\begin{pmatrix}1 & x_j\\ x_j & x_j^2\end{pmatrix}. \qquad 4.2.1\text{-}10$$

These equations were evaluated with the FORTRAN subroutine, PSNERR.FOR, which is listed in Appendix 2, Section A2-2.2. The maximum likelihood estimates were calculated by a Newton-Raphson algorithm which was implemented in the Gauss programming language; MLE.GSS, Section A2-2.1.

The maximum likelihood estimator involves considerable calculation and a typical estimation at $n = 1600$ took 10 to 15 minutes on a 486/50 personal computer. Thus the computing time that was required to complete each series in Table 2 was about one week. This time requirement was the primary consideration in limiting the largest sample size to 1600 and the number of data sets per sample size to 500.

4.2.2 Iteratively Reweighted Least Squares Estimates

The iteratively reweighted least squares estimates for the quasi-structural model follow from direct substitution of $\mathrm{F}_X^n(\cdot)$ in Equations 2.2.2-8 and 2.2.2-9. This

gives $$E_{Y|Z=z}(y) = \alpha + \beta\, \xi_{F_X^n}(z) \qquad 4.2.2\text{-}1$$

and $$V_{Y|Z=z}(y) = \alpha + \beta\, \xi_{F_X^n}(z) + \beta^2\, V_{F_X^n}(z) \qquad 4.2.2\text{-}2$$

where $$\xi_{F_X^n}(z_i) = \sum_{j=1}^{n} x_j\, c_{ij}/c_{i\cdot} \qquad 4.2.2\text{-}3$$

and $$V_{F_X^n}(z_i) = \sum_{j=1}^{n} x_j^2\, c_{ij}/c_{i\cdot} - \left[\xi_{F_X^n}(z_i)\right]^2 . \qquad 4.2.2\text{-}4$$

Both $\xi_{F_X^n}(z)$ and $V_{F_X^n}(z)$ were calculated by the FORTRAN subroutine, EVX_Z.FOR, Appendix 2, Section A2-2.4. The actual estimates of α and β were evaluated using a simple iteratively reweighted least squares algorithm that was programmed as the Gauss procedure, WLS_PSN, Section A2-2.3.

4.2.3 Replacement Estimator

The replacement estimator can be calculated in the same way as the IRLS estimator. The distinction between the estimators is that the replacement estimator omits the final term of Equation 4.4.2-2 to give

$$V_{Y|Z=z}(y) = \alpha + \beta\, \xi_{F_X^n}(z). \qquad 4.2.3\text{-}1$$

Both the IRLS estimator and the replacement estimator were calculated in the Gauss program, EVX_Z.GSS, Appendix 2, Section A2-2.5.

4.3 IRLS Estimator for the Structural Model

Since $F_X(\cdot)$ is known in the simulations, the estimators for a structural model are defined by the equations of Section 2.2 just as they were defined for the quasi-structural model. The maximum likelihood estimator in this measurement error model requires numerical integration. However, the iteratively reweighted least squares estimates have a closed form and they are relatively easy to calculate. These estimates were produced primarily to demonstrate that the quasi-structural model is at least as efficient as the structural model.

As shown in the previous section, the estimates for α and β can be calculated using the WLS_PSN program (Appendix 2, Section A2-2.3). This procedure requires $\xi_{F_X}(z)$ and $V_{F_X}(z)$ as input.

For any mixture distribution, $G(x) = pG_1(x) + (1-p)G_2(x)$, and error model, $f_{Z|X}(\cdot)$, $f_{X|Z=z}(x;G)$ can be calculated as

$$f_{X|Z=z}(x;G) \propto p\, f_Z(z;G_1) f_{X|Z=z}(x;G_1) + (1-p) f_Z(z;G_2) f_{X|Z=z}(x;G_2) \qquad 4.3\text{-}1$$

where

$$f_Z(z;G_k) = \int f_{X|Z=z}(x;G_k)\, dG_k(x)\,. \qquad 4.3\text{-}2$$

Then $\xi_G(z)$ and $V_G(z)$ can be written as

$$f_Z(z;G)\xi_G(z) = p\, f_Z(z;G_1)\xi_{G_1}(z) + (1-p) f_Z(z;G_2)\xi_{G_2}(z) \qquad 4.3\text{-}3$$

and

$$f_Z(z;G)V_G(z) = p\, f_Z(z;G_1)V_{G_1}(z) + (1-p) f_Z(z;G_2)V_{G_2}(z). \qquad 4.3\text{-}4$$

Thus, working with any mixture is about as tractable as working with the component distributions. In the case where $f_{Z|X}(\cdot)$ and $G_k(\cdot)$ are lognormal distributions, the components can be expressed in closed form which speeds up the computations as well as program development.

Let $f_{Z|X}(\cdot)$ be the lognormal distribution with parameters,

$\mu = \ln(x) - \ln\left(\sqrt{\nu^2+1}\right)$ and $\sigma^2 = \ln\left(\nu^2+1\right)$. These parameter choices will give a

distribution that has a constant coefficient of variation, ν , and expectation, x . Also let $G_k(\cdot)$ be a lognormal distribution with parameters, θ_k and ω_k^2 , then $f_{X|Z=z}(x;G_k)$ and $f_Z(z;G_k)$ will be lognormal distributions. If $f_{X|Z=z}(x;G_k)$ is parameterized with $\theta(z)$ and ω^2, then

$$\theta(z)=\frac{\theta_k \ln(\nu^2+1)+\omega_k^2\left(\ln(z)+\ln\left(\sqrt{\nu^2+1}\right)\right)}{\ln(\nu^2+1)+\omega_k^2} \qquad 4.3\text{-}5$$

and

$$\omega^2=\frac{\omega_k^2 \ln(\nu^2+1)}{\ln(\nu^2+1)+\omega_k^2}\,. \qquad 4.3\text{-}6$$

From this parameterization, the mean and variance are obtained as

$$\xi_{G_k}(z)=\exp\left[\theta(z)+\tfrac{1}{2}\omega^2\right] \qquad 4.3\text{-}7$$

and

$$V_{G_k}(z)=\exp\left[2(\theta(z)+\omega^2)\right]-\left[\xi_{G_k}(z)\right]^2. \qquad 4.3\text{-}8$$

The distribution of $f_Z(z;G_k)$ has the parameters, $\theta_k-\ln\left(\sqrt{\nu^2+1}\right)$ and $\ln(\nu^2+1)+\omega_k^2$. These results are more completely verified in Appendix 1, A1-5.

The computer program, MIXTURE.GSS, produced the estimates under the mixture model for the distribution of the unobserved covariate. This program is given in Appendix 2, Section A2-3.1. The computer program, LN_NORM.GSS, produced the estimates under the lognormal model for the distribution of the unobserved covariates. This program is given in Section A2-3.2.

4.4 Correction for Attenuation Method

There is no widely recognized correction for attenuation estimator under this measurement error model largely because the necessary correction depends upon the regression parameters and the unobserved covariates and thus they are unknown. Nevertheless, as discussed in Section 2.5, improved estimates can be calculated simply by using the naive estimates and $\mathbf{z}$ to replace the unknown quantities. Then a "correction for attenuation" is given by Equation 2.5-9.

The procedure, as implemented, calculates the naive estimates. That is, $\left(\overset{*}{\hat{\alpha}}_{(0)}, \overset{*}{\hat{\beta}}_{(0)}\right)$ satisfying

$$\overset{*}{\mathbf{U}}(\alpha, \beta; \mathbf{y}, \mathbf{z}) = \mathbf{0} \ . \tag{4.4-1}$$

Because the naive estimates can be very biased, the estimator can be improved by taking the adjustment through a few iterations. That is, let

$$W_i = \frac{1}{\overset{*}{\hat{\alpha}}_{(k)} + \overset{*}{\hat{\beta}}_{(k)} z_i} \tag{4.4-2}$$

and

$$B_{(k)} = \frac{\sum_{i=1}^{n} W_i (z_i - \bar{z})^2}{\sum_{i=1}^{n} W_i (z_i - \bar{z})^2 - \sum_{i=1}^{n} W_i \sigma_i^2} \tag{4.4-3}$$

where

$$\sigma_i^2 = \frac{\nu^2 z_i^2}{\nu^2 + 1} \ . \tag{4.4-4}$$

Then the "correction for attenuation" estimators were calculated as

$$\overset{*}{\hat{\beta}}_{(k+1)} = B_{(k)} \overset{*}{\hat{\beta}}_{(0)} \tag{4.4-5}$$

and

$$\hat{\alpha}^{*}_{(k)} = \frac{\sum_{i=1}^{n} W_i \left(y_i - \hat{\beta}^{*}_{(k)} z_i \right)}{\sum_{i=1}^{n} W_i} . \qquad 4.5\text{-}6$$

Conceptually one could stop with $\left(\hat{\alpha}^{*}_{(1)}, \hat{\beta}^{*}_{(1)} \right)$ and satisfy the spirit of the correction for attenuation. However, after a few iterations $B_{(k)}$ stabilizes at a value which simulation experience indicates to be slightly better. The computer program is reproduced in Appendix 2, Section A2-4.1. This program, ATTN.GSS, calculated $\left(\hat{\alpha}^{*}_{(5)}, \hat{\beta}^{*}_{(5)} \right)$. Since there is little change in these estimates after a few iterations, this convention simply avoided the added programming that would be required to enforce a more sophisticated stopping rule.

The expression in Equation 4.4-4 has the variance, $(vx_i)^2$, for its expected value. When the weights are assumed known, this choice in the expression, $\sum_{i=1}^{n} W_i \sigma_i^2$, yields a consistent estimate of $\sum_{i=1}^{n} W_i (vx_i)^2$. A more detailed derivation of this is contained in Appendix 1, Section A1-2 and in Hwang (1986).

4.5 Approximate Nakamura Method

The technique that was outlined in Section 2.6 was applied to the simulated measurement error model with multiplicative errors. Even through a Taylor's approximation to the corrected score equations was necessary, the simulations indicate that the Nakamura method gave essentially unbiased estimates. This consistency is established using the moments of $f_{Z|X}(\cdot)$.

The naive score equations for this Poisson model are the sum of the terms,

$$\overset{*}{\mathbf{U}}_i(\alpha,\beta;y_i,\mathrm{x}) = \left(y_i/(\alpha+\beta\mathrm{x})-1\right)\begin{bmatrix}1\\ \mathrm{x}\end{bmatrix} \qquad 4.5\text{-}1$$

The derivatives of Equations 4.5-1 are

$$\frac{\partial^k}{\partial \mathrm{x}^k}\overset{*}{\mathbf{U}}_i(\alpha,\beta;y_i,\mathrm{x})\Big/k! = \left(\frac{(-\beta)^{k-1}y}{(\alpha+\beta\mathrm{x})^{k+1}}\right)\begin{bmatrix}-\beta\\ \alpha\end{bmatrix} \quad ; \quad k\geq 2\,. \qquad 4.5\text{-}2$$

This formula for the derivative is easily verified by the method of induction.

As discussed in Section 2.6, Nakamura estimators can be obtained by expanding the naive score equations about the true covariate values and then taking the expectation with respect to the distribution, $f_{Z|X}(\cdot)$. In this case, taking the expectation of the Taylor's expansion of the naive score, $\overset{*}{\mathbf{U}}(\alpha,\beta;\mathbf{y},\mathbf{z})$, about $\mathbf{z}=\mathbf{x}$ results in the equations,

$$\mathrm{E}_{Z|X}\left[\overset{*}{\mathbf{U}}(\alpha,\beta;\mathbf{y},\mathbf{z})\right] = \begin{bmatrix} \sum_{i=1}^{n}\left(\frac{y_i}{\alpha+\beta x_i}-1\right)+\sum_{i=1}^{n}y_i\left[\sum_{k=2}^{\infty}(-\beta)^k\,\frac{\mathrm{E}_{Z|X}(x_i-\mathrm{z})^k}{(\alpha+\beta x_i)^{k+1}}\right] \\ \sum_{i=1}^{n}x_i\left(\frac{y_i}{\alpha+\beta x_i}-1\right)+\sum_{i=1}^{n}\alpha\, y_i\left[\sum_{k=2}^{\infty}(-\beta)^{k-1}\,\frac{\mathrm{E}_{Z|X}(x_i-\mathrm{z})^k}{(\alpha+\beta x_i)^{k+1}}\right] \end{bmatrix}. \qquad 4.5\text{-}3$$

In the notation of Section 2.6, this gives

$$\mathbf{T}_k(\mathbf{x}) = \begin{bmatrix} \sum_{i=1}^{n} y_i \, (-\beta)^k \, \dfrac{\mathrm{E}_{Z|X}(x_i - \mathrm{z})^k}{(\alpha + \beta x_i)^{k+1}} \\ \sum_{i=1}^{n} \alpha \, y_i \, (-\beta)^{k-1} \, \dfrac{\mathrm{E}_{Z|X}(x_i - \mathrm{z})^k}{(\alpha + \beta x_i)^{k+1}} \end{bmatrix}. \qquad \text{4.5-4}$$

To my knowledge there is not a simple function of $\mathbf{z}$ with expectation, $\sum_{k=2}^{\infty} \mathbf{T}_k(\mathbf{x})$, so a consistent estimate, $\mathbf{t}_2(\mathbf{z})$ was used for the $k = 2$ term. The crux of the Nakamura method is to replace x^k with $c_k \, z^k$ where c_k is chosen to make $c_k \, \mathrm{E}_{Z|X}(\mathrm{z}^k) = x^k$. It is easy to derive the required constants for the lognormal error distribution, $\mathrm{f}_{Z|X}(\cdot)$, and the expressions for $\mathrm{E}_{Z|X}(x_i - \mathrm{z})^k$ and $(\alpha + \beta x_i)^k$ can be expanded in powers of x to accommodate the substitution. Using this replacement and truncating the Taylor's expansion at $k = 2$ yields the correction to the score equations,

$$\mathbf{t}_2(\mathbf{z}) = \begin{bmatrix} \sum_{i=1}^{n} \dfrac{-y_i \, \beta^2 \, \nu^2 \, z_i^2 / (\nu^2 + 1)}{\alpha^3 + 3\alpha^2 \beta \, z_i + 3\alpha \, \beta^2 \, z_i^2 / (\nu^2 + 1) + \beta^3 \, z_i^3 / (\nu^2 + 1)^3} \\ \sum_{i=1}^{n} \dfrac{\alpha \, \beta \, y_i \, \nu^2 \, z_i^2 / (\nu^2 + 1)}{\alpha^3 + 3\alpha^2 \beta \, z_i + 3\alpha \, \beta^2 \, z_i^2 / (\nu^2 + 1) + \beta^3 \, z_i^3 / (\nu^2 + 1)^3} \end{bmatrix}. \qquad \text{4.5-5}$$

The solution, $\hat{\alpha}$ and $\hat{\beta}$, for the equations, $\overset{*\,C}{\mathbf{U}}(\alpha, \beta; \mathbf{y}, \mathbf{z}) = \mathbf{0}$, was obtained using a Newton-Raphson algorithm (NAK2MURA.GSS, Appendix 2, Section A2-5.1). This algorithm requires the derivatives of $\overset{*\,C}{\mathbf{U}}(\alpha, \beta; \mathbf{y}, \mathbf{z})$. Because the Newton-

Raphson algorithm is not very sensitive to the choice and because the estimated standard errors for the regression estimates depends upon this matrix, the connection to the quasi-structural model was used. Thus Equations 4.5-5 and 4.5-6 were truncated at $k = 2$ and their derivatives were taken with respect to the parameters to give the set of equations,

4.5-9

$$\mathrm{E}_{Z|X}\left(\frac{\partial \overset{*}{\mathrm{U}}_{\alpha}(\alpha,\beta;\mathbf{y},\mathbf{z})}{\partial\alpha}\right) \approx \sum_{i=1}^{n}\left(\frac{-y_i}{(\alpha+\beta x_i)^2} + \frac{3\beta y_i \mathrm{E}_{Z|X}(x_i-\mathrm{z})^2}{(\alpha+\beta x_i)^4}\right)$$

4.5-10

$$\mathrm{E}_{Z|X}\left(\frac{\partial \overset{*}{\mathrm{U}}_{\alpha}(\alpha,\beta;\mathbf{y},\mathbf{z})}{\partial\beta}\right) \approx \sum_{i=1}^{n}\left(\frac{-x_i y_i}{(\alpha+\beta x_i)^2} - \frac{\beta y_i(\beta x_i - 2\alpha)\mathrm{E}_{Z|X}(x_i-\mathrm{z})^2}{(\alpha+\beta x_i)^4}\right)$$

4.5-11

$$\mathrm{E}_{Z|X}\left(\frac{\partial \overset{*}{\mathrm{U}}_{\beta}(\alpha,\beta;\mathbf{y},\mathbf{z})}{\partial\beta}\right) \approx \sum_{i=1}^{n}\left(\frac{-x_i^2 y_i}{(\alpha+\beta x_i)^2} - \frac{\alpha\beta y_i(\beta x_i - 2\alpha)\mathrm{E}_{Z|X}(x_i-\mathrm{z})^2}{(\alpha+\beta x_i)^4}\right)$$

Once again each occurrence of x^k is replaced with $c_k z^k$ to yield the equations programmed in Section A2-5.1. The procedure as presented uses the first four moments of the distribution, $\mathrm{f}_{Z|X}(\cdot)$. The sensitivity of this procedure to moments based on misspecification of $\mathrm{f}_{Z|X}(\cdot)$ was not examined.

4.6 Replacement Methods

Replacement methods are very direct to set up in their rudimentary version since $\xi_G(z_i)$ simply replaces z_i within the naive likelihood equations. For the simulated models, the iteratively reweighted least squares algorithm, WLS_PSN (Appendix 2, Section A2-2.3), can calculate the parameter estimates in these cases by setting $V_G(\mathbf{z}) = \mathbf{0}$. Under an assumed $G(\cdot)$, the iteratively reweighted least squares estimates for the resulting structural model apply this algorithm with

$$E_{Y|Z=z_i}(y) = \alpha + \beta\, \xi_G(z_i) \qquad 4.6\text{-}1$$

and

$$V_{Y|Z=z_i}(y) = \alpha + \beta\, \xi_G(z_i) + \beta^2\, V_G(z_i)\ . \qquad 4.6\text{-}2$$

So the distinction between these two models is simply the term, $\beta^2\, V_G(z_i)$. Pierce *et al.* (1992) noted that this term was negligible in cancer modeling for the Life Span Study. In the models that were simulated here, this was true with respect to the efficiency of estimates for α and β but including $\beta^2\, V_G(z_i)$ significantly improved the variance estimates of these regression parameters.

Two replacement methods were run in the simulations. The Schafer method does not have a natural estimate for $V_G(z_i)$ in as much as Schafer (1990) discourages using a linear or quadratic empirical Bayes estimator for this term. However, the modified Whittemore method has a simple and natural estimate for $V_G(z_i)$. For this reason, this correction was applied in programming the Whittemore method and it was ignored in programming the Schafer method. Both of these methods, along with the Poisson regression using $\mathbf{x}$ and the naive estimates, were calculated by the program, WLS_SIM, which is reproduced in Appendix 2, Section A2-6.1.

4.6.1 Modified Whittemore Method

The modified Whittemore method has been explained in Section 2.7.1. As programmed for the simulations, $\xi_G(z_i)$ and $V_G(z_i)$ were calculated as defined in Equations 2.7.1-6 and 2.7.1-7.

4.6.2 Schafer Method.

The quadratic estimator of Schafer (1990) was calculated in the simulations. The general procedure was discussed in Section 2.7.2. For a polynomial in z approximation to $\xi_G(z)$ of degree q, Equation 2.7.2-4 requires the evaluation of the terms,

$$\int x\, E_{Z|X}(z^r)\, dG(x);\ r = 0,\ 1,\ \cdots\ q\ . \qquad 4.6.2\text{-}1$$

Under the lognormal error model as specified here, Equations 4.1-3, 4.1-4, and 4.1-5 imply that

$$E_{Z|X=x}(z^r) = x^r\left(\nu^2+1\right)^{(r^2-r)/2} \qquad 4.6.2\text{-}2$$

so that

$$\int x\, E_{Z|X}(z^r)\, dG(x) = \left(\nu^2+1\right)^{(r^2-r)/2} E_X\left(x^{r+1}\right)\ . \qquad 4.6.2\text{-}3$$

Noting that

$$E_Z(z^k) = E_X\left[E_{Z|X=x}(z^k)\right] = E_X(x^k)\left(\nu^2+1\right)^{(k^2-k)/2} \qquad 4.6.2\text{-}4$$

gives

$$\int x\, E_{Z|X}(z^r)\, dG(x) = \frac{E_Z(z^{r+1})}{\left(\nu^2+1\right)^r}\ . \qquad 4.6.2\text{-}5$$

This gives a system of $q+1$ equations that depend upon the coefficients, θ, the moments of the observed covariate, $E_Z(z^r)$, and the known quantities, $\left(\nu^2+1\right)^r$. Substituting

$$\frac{\sum_{i=1}^{n} z_i^r}{n} \qquad 4.6.2\text{-}6$$

for $E_z(z^r)$ gives a system of equations which can be represented as

$$\mathbf{D}'\mathbf{D}\theta = \mathbf{b} \ , \qquad 4.6.2\text{-}7$$

where the "i"th row of $\mathbf{D}$ is $(1\, z_i\, z_i^2 \cdots z_i^q)$ and the "r"th row of $\mathbf{b}$ is

$$\frac{\sum_{i=1}^{n} z_i^r}{(v^2+1)^{r-1}} \ . \qquad 4.6.2\text{-}8$$

Equation 4.6.2-7 has the solution, $\hat{\theta} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{b}$ and the method replaces $\mathbf{z}$ with $\xi_q(\mathbf{z}) = \mathbf{D}\hat{\theta}$.

Note to reader: The thesis conclusions are given in Section 3.4.

Bibliography

Amemiya, Y and W A Fuller. 1988. Estimation for the nonlinear functional relationship. The Annals of Statistics 16(1):147-160.

Armstrong, B. 1985. Measurement error in the generalized linear model. Communications in Statistics - Simulation and Computation 14(3):529-544.

Burr, D. 1988. On errors-in-variables in binary regression - Berkson Case. Journal of the American Statistical Association 83:739-743.

Carroll, R J, C H Spiegelman, K K G Lan, K T Bailey and R D Abbott. 1984. On errors-in-variables for binary regression models. Biometrika 71(1):19-25.

Cochran, W G. 1968. Errors of measurement in statistics. Technometrics 10:55-83.

Fishman, C S. 1976. Sampling from the Poisson distribution on a computer. Computing 17:145-156.

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

Gilbert, E S. 1984. Some effects of random dose measurement errors on the analyses of Atomic Bomb survivor data. Radiation Research 98:591-605.

Gong, G and F J Samaniego. 1981. Pseudo maximum likelihood estimation: theory and applications. The Annals of Statistics 9(4):861-869.

Hsiao, C. 1989. Consistent estimation for some nonlinear errors-in-variables models. Journal of Econometrics 41:159-185.

Hwang, J T. 1986. Multiplicative errors-in-variables models with application to recent data released by the US Department of Energy. Journal of the American Statistical Association 81:680-688.

Jablon, S. 1971. Atomic Bomb radiation dose estimation at ABCC. Atomic Bomb Casualty Commission TR 23-71. Hiroshima: Radiation Effects Research Foundation.

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

Maritz,.J S and T Lwin. 1989. *Empirical Bayes Methods* (2nd. ed.). New York: Chapman and Hall.

McCullagh, P and J A Nelder. 1989. *Generalized Linear Models* (2nd. ed.). New York: Chapman and Hall.

Nakamura, T. 1990. Corrected score functions for errors-in-variables models: Methodology and application to generalized linear model. Biometrika 77(1):127-137.

Nakamura, T. 1992. Proportional hazards model with covariates subject to measurement error. Biometrics 48:829-838.

Pierce, D A, D O Stram and M Vaeth. 1990. Allowing for random errors in radiation dose estimates for the atomic bomb survivor data. Radiation Research 123:275-284.

Pierce, D A, D O Stram, M Vaeth and D W Schafer. 1992. The errors-in-variables problem: Considerations provided by radiation dose-response analyses of the A-bomb survivor data. Journal of the American Statistical Association 87(418):351-359.

Prentice, R L. 1982. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69:331-342.

Schafer, D W. 1990. Measurement error model estimation using iteratively weighted least squares. Contemporary Mathematics 112:129-138.

Schafer, D W. 1992. Replacement methods for measurement error models. Technical Report 151, Department of Statistics, Oregon State University. 15 pp.

Stefanski, L A and R J Carroll. 1985. Covariate measurement error in logistic regression. The Annals of Statistics 13(4):1335-1351.

Stefanski, L A and R J Carroll. 1987. Conditional scores and optimal scores for generalized linear measurement-error models. Biometrika 74(4):703-716.

Stefanski, L A and R J Carroll. 1990. Score tests in generalized linear measurement error models. Journal of the Royal Statistical Society, Series B 52(2):345-359.

Stefanski, L A and R J Carroll. 1991. Deconvolution-based score tests in measurement error models. The Annals of Statistics 19(1):249-259.

Whittemore, A S. 1989. Errors-in-variables regression using Stein estimates. The American Statistician 43(4):226-228

Appendices

Appendix 1

A1-1 Likelihood Score and Observed Information in a Structural Measurement Error Model

This appendix contains algebraic details of the derivation of the score equation and observed information as discussed in Section 2.2.1. Both results follow directly from the definitions of the likelihood score and the observed information under the distributional assumptions of Section 2.1. Two of these assumptions that are expressly made in these derivations are:

1. Given x, y and z are independent,
2. and $f_{Z|X}(\cdot)$ is not a function of φ.

A1-1.1 Verification of the Score, Equation 2.2.1-5

2.2.1-5

$$U_r(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) = \sum_{i=1}^{n} \int \overset{*}{U}_{ri}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X)\, dx$$

Proof: From the definition of the likelihood score as the derivitive of the log-likelihood, the components of the score are:

$$U_r(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) = \sum_{i=1}^{n} \frac{\partial}{\partial \varphi_r} \ln\left[\int \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx\right]$$

$$= \sum_{i=1}^{n}\left[\frac{\int \frac{\partial}{\partial \varphi_r}\left[\overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x)\right] f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx}{\int \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx}\right]$$

$$= \sum_{i=1}^{n}\left[\frac{\int \frac{\partial}{\partial \varphi_r}\left[\ln \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x)\right] \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx}{\int \overset{*}{L}_i(\varphi; y_i, \mathbf{u}_i, x) f_{X|UZ}(x|\mathbf{u}_i, z_i; F_X)\, dx}\right].$$

Substituting $f_Y(y_i;g(\varphi'\mathbf{w}_i))$ for $\overset{*}{L}_i(\varphi;y_i,\mathbf{u}_i,x)$ gives

$$U_r(\varphi;F_X,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\left[\frac{\int\frac{\partial}{\partial\varphi_r}\left[\ln\overset{*}{L}_i(\varphi;y_i,\mathbf{u}_i,x)\right]f_Y(y_i;g(\varphi'\mathbf{w}_i))f_{X|UZ}(x|\mathbf{u}_i,z_i;F_X)dx}{\int\overset{*}{L}_i(\varphi;y_i,\mathbf{u}_i,x)f_{X|UZ}(x|\mathbf{u}_i,z_i;F_X)dx}\right]$$

and noting that

$$f_{X|YUZ}(x|y,\mathbf{u},z;F_X)\propto f_Y(y_i;g(\varphi'\mathbf{w}))f_{X|UZ}(x|\mathbf{u},z;F_X)$$

gives

$$U_r(\varphi;F_X,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\int\frac{\partial}{\partial\varphi_r}\left[\ln\overset{*}{L}_i(\varphi;y_i,\mathbf{u}_i,x)\right]f_{X|YUZ}(x|y_i,\mathbf{u}_i,z_i;F_X)dx$$

Denoting the naive score as

$$\overset{*}{U}_{ri}(\varphi;y_i,\mathbf{u}_i,x)\equiv\frac{\partial}{\partial\varphi_r}\left[\ln\overset{*}{L}_i(\varphi;y_i,\mathbf{u}_i,x)\right]$$

completes the proof, giving

$$U_r(\varphi;F_X,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\int\overset{*}{U}_{ri}(\varphi;y_i,\mathbf{u}_i,x)f_{X|YUZ}(x|y_i,\mathbf{u}_i,z_i;F_X)dx\ .$$

A1-1.2 Verification of the Observed Information, Equation 2.2.1-7:

2.2.1-7

$$\mathrm{i}_{\mathrm{rc}}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\int \overset{*}{\mathrm{i}}_{\mathrm{rc}i}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}$$

$$-\sum_{i=1}^{n}\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\overset{*}{\mathrm{U}}_{ci}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}$$

$$+\sum_{i=1}^{n}\left[\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\quad\int \overset{*}{\mathrm{U}}_{ci}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right]$$

Proof: By definition, the observed information matrix has elements,

$$\mathrm{i}_{\mathrm{rc}}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z})=-\sum_{i=1}^{n}\frac{\partial}{\partial\varphi_{\mathrm{c}}}\left[\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right].$$

Then

$$\mathrm{i}_{\mathrm{rc}}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z})=-\sum_{i=1}^{n}\left\{\int \frac{\partial}{\partial\varphi_{\mathrm{c}}}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\right]\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right\}$$

$$-\sum_{i=1}^{n}\left\{\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\frac{\partial}{\partial\varphi_{\mathrm{c}}}\left[\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\right]\mathrm{dx}\right\}.$$

Defining

$$\overset{*}{\mathrm{i}}_{\mathrm{rc}i}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\equiv-\frac{\partial}{\partial\varphi_{\mathrm{c}}}\left[\overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\right],$$

gives

$$\mathrm{i}_{\mathrm{rc}}(\boldsymbol{\varphi};\mathrm{F}_X,\mathbf{y},\mathbf{u},\mathbf{z})=\sum_{i=1}^{n}\left\{\int \overset{*}{\mathrm{i}}_{\mathrm{rc}i}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\,\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\,\mathrm{dx}\right\}$$

$$-\sum_{i=1}^{n}\left\{\int \overset{*}{\mathrm{U}}_{ri}(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x})\frac{\partial}{\partial\varphi_{\mathrm{c}}}\left[\mathrm{f}_{X|YUZ}(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X)\right]\mathrm{dx}\right\}.$$

Considering further the term,

$$\frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)\right],$$

where

$$\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)=\frac{\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)}{\int \mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}},$$

yields

$$\frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)\right]=\frac{\frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\right]\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)}{\int \mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}}$$

$$-\left[\frac{\int \frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\right]\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}}{\int \mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}}\right]\left[\frac{\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)}{\int \mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\mathrm{f}_{X|UZ}\left(\mathrm{x}|\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}}\right].$$

Noting that

$$\frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right)\right]=\overset{*}{\mathrm{U}}_{ci}\left(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x}\right)\mathrm{f}_Y\left(y_i;\mathrm{g}(\boldsymbol{\varphi}'\mathbf{w}_i)\right),$$

gives

$$\frac{\partial}{\partial \varphi_c}\left[\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)\right]=\overset{*}{\mathrm{U}}_{ci}\left(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x}\right)\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)$$

$$-\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)\int \overset{*}{\mathrm{U}}_{ci}\left(\boldsymbol{\varphi};y_i,\mathbf{u}_i,\mathrm{x}\right)\mathrm{f}_{X|YUZ}\left(\mathrm{x}|y_i,\mathbf{u}_i,z_i;\mathrm{F}_X\right)\mathrm{dx}.$$

This essentially completes the derivation since substitution of this expression into the previous result for $i_{rc}(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z})$ gives

$$i_{rc}(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) = \sum_{i=1}^{n} \left\{ \int \overset{*}{i}_{rci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx \right\}$$

$$- \sum_{i=1}^{n} \left\{ \int \overset{*}{U}_{ri}(\varphi; y_i, \mathbf{u}_i, x) \left[\overset{*}{U}_{ci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) \right] dx \right\}$$

$$- \sum_{i=1}^{n} \left\{ \int \overset{*}{U}_{ri}(\varphi; y_i, \mathbf{u}_i, x) \left[-f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) \int \overset{*}{U}_{ci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx \right] dx \right\}$$

which yields the desired equation for the observed information,

$$i_{rc}(\varphi; F_X, \mathbf{y}, \mathbf{u}, \mathbf{z}) = \sum_{i=1}^{n} \int \overset{*}{i}_{rci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx$$

$$- \sum_{i=1}^{n} \int \overset{*}{U}_{ri}(\varphi; y_i, \mathbf{u}_i, x) \overset{*}{U}_{ci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx$$

$$+ \sum_{i=1}^{n} \left[\int \overset{*}{U}_{ri}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx \int \overset{*}{U}_{ci}(\varphi; y_i, \mathbf{u}_i, x) f_{X|YUZ}(x|y_i, \mathbf{u}_i, z_i; F_X) dx \right].$$

A1-2 Three Examples of a Technique to Obtain Corrected Score Equations

A technique that can be used to find corrected score equations (Nakamura, 1990) was outlined in Section 2.6. In Section 4.5 this technique was applied to a Poisson model with multiplicative measurement errors, a model that is unmanageable using other methods. Both the general outline and the example camouflage the uncomplicated derivations that can accompany the technique. It should be noted that general functional approaches tend to be complicated and the example in Section 4.5 is algebraically difficult. When applied in measurement error models where corrections for attenuation are known, this technique is elegant. The usual normal theory correction for attenuation estimate for additive errors (Fuller, 1987) is easily derived and the extention to multiplicative errors is almost trivial whereas more traditional derivations (Hwang, 1986) are somewhat cumbersome. The multiplicative extention also gives an elementary example of finding a function of the observed data, $\mathbf{t}_2(\mathbf{z})$, such that $\mathrm{E}_{Z|X=x}(\mathbf{t}_2(\mathbf{z})) = \mathbf{T}_2(\mathbf{x})$.

The final example, a log-linear Poisson model with additive normally distributed measurement errors, illustrates the tractability of the technique in dealing with nonlinear parameterizations and weights that depend upon the parameters. The corrected score equations for this example were given by Nakamura (1990). However, his derivation employs a more direct technique that is quite difficult to implement in circumstances where the requisite expectations are unknown, such as the case discussed in Section 4.5.

In the normal theory setting presented in Section 2.5, it was assumed that $\mathrm{E}_{Z|X=x}(\mathrm{z}) = x$ and $\mathrm{V}_{Z|X=x}(\mathrm{z}) = \sigma^2$. As presented there, the corrected for attenuation estimate of the slope is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i (z_i - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2 - n\sigma^2}.$$

Note that this estimator is equivalent to the β that satisfies the corrected score equations;

$$\overset{*}{\mathbf{U}}{}^{C}(\alpha,\beta;\mathbf{y},\mathbf{z})=\begin{bmatrix}\sum_{i=1}^{n}(y_i-\alpha-\beta z_i)\\ \sum_{i=1}^{n}(y_i-\alpha-\beta z_i)\,z_i+n\beta\sigma^2\end{bmatrix}=\mathbf{0}\ .$$

These equations follow directly from Equation 2.6-5 since

$$\frac{\partial^2}{\partial x_i^2}\left(\overset{*}{\mathbf{U}}_i(\alpha,\beta;y_i,x_i)\right)=\begin{bmatrix}0\\-2\beta\end{bmatrix}$$

implies that

$$\mathbf{t}_2(\mathbf{z})=\mathbf{T}_2(\mathbf{x})=\begin{bmatrix}\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}\left(\overset{*}{\mathrm{U}}_{\alpha i}(\alpha,\beta;y_i,x_i)\right)\frac{\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^2}{2}\\ \sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}\left(\overset{*}{\mathrm{U}}_{\beta i}(\alpha,\beta;y_i,x_i)\right)\frac{\mathrm{E}_{Z|X=x_i}(\mathrm{z}-x_i)^2}{2}\end{bmatrix}=\begin{pmatrix}0\\-n\beta\sigma^2\end{pmatrix}$$

and

$$\mathbf{t}_k(\mathbf{z})=\begin{pmatrix}0\\0\end{pmatrix};\ k\neq 2\ .$$

The multiplicative error model essentially differs from the additive model in that $\mathrm{V}_{Z|X=x}(\mathrm{z})=\nu^2x^2$. Replacing σ^2 in the previous model with ν^2x^2 gives

$$\mathbf{T}_2(\mathbf{x})=\begin{pmatrix}0\\-\beta\nu^2\sum_{i=1}^{n}x_i^2\end{pmatrix}.$$

Then noting that $\mathrm{E}_{Z|X=x}(\mathrm{z}^2)=\nu^2x^2+x^2$ suggests taking

$$\mathbf{t}_2(\mathbf{z}) = \begin{pmatrix} 0 \\ -\beta \dfrac{\nu^2}{\nu^2+1} \sum_{i=1}^{n} z_i^2 \end{pmatrix}$$

so that

$$\mathbf{E}_{Z|X=x}\left(\mathbf{t}_2(\mathbf{z})\right) = \mathbf{T}_2(\mathbf{x}) \ .$$

This yields the corrected score equations,

$$\dot{\mathbf{U}}^{C}(\alpha,\beta;\mathbf{y},\mathbf{z}) = \begin{bmatrix} \sum_{i=1}^{n}(y_i - \alpha - \beta z_i) \\ \sum_{i=1}^{n}(y_i - \alpha - \beta z_i)\, z_i + \beta \dfrac{\nu^2}{\nu^2+1} \sum_{i=1}^{n} z_i^2 \end{bmatrix} = \mathbf{0} \ ,$$

which duplicates the result of Hwang (1986).

The Poisson model with $\mu = \exp(\alpha + \beta x)$ and normally distributed measurement errors, $z \sim \phi\left((z-x)/\sigma\right)$, has been shown by Nakamura (1990) and Armstrong (1985) to have a nice solution. For this Poisson model, the naive score equations are the sum of terms,

$$\begin{bmatrix} \overset{*}{U}_{\alpha i}(\alpha,\beta;y_i,x_i) \\ \overset{*}{U}_{\beta i}(\alpha,\beta;y_i,x_i) \end{bmatrix} = \begin{bmatrix} y_i - e^{\alpha+\beta x_i} \\ \left(y_i - e^{\alpha+\beta x_i}\right) x_i \end{bmatrix},$$

and the corrected score equations,

$$\overset{*\,C}{\mathbf{U}}(\alpha,\beta;\mathbf{y},\mathbf{z})=\begin{bmatrix}\sum_{i=1}^{n}\left(y_i-e^{\alpha+\beta z_i-\beta^2\sigma^2/2}\right)\\ \sum_{i=1}^{n}\left(y_i-e^{\alpha+\beta z_i-\beta^2\sigma^2/2}\right)z_i\end{bmatrix}=\mathbf{0}\ ,$$

are straightforward to derive by noting that $E_{Z|X=x}\left[e^{\beta z}\right]$ is the moment generating function for the normal distribution, $\phi\left((z-x)/\sigma\right)$. Hence,

$$E_{Z|X=x}\left[e^{\beta z}\right]=e^{\beta x+\beta^2\sigma^2/2}\ \text{ and }\ E_{Z|X=x}\left[z\,e^{\beta z}\right]=x\,e^{\beta x+\beta^2\sigma^2/2}.$$

The need to recognize the required expectations limits the application of the Nakamura method to a few special cases. To appreciate the limitation simply consider the difficulty introduced by a multiplicative error, i.e. $\sigma^2=\nu^2x_i^2$, yielding

$$E_{Z|X=x}\left[e^{\beta z}\right]=e^{\beta x+\beta^2\nu^2x^2/2}\ \text{ and }\ E_{Z|X=x}\left[z\,e^{\beta z}\right]=x\,e^{\beta x+\beta^2\nu^2x^2/2}\ .$$

The techique of Section 2.6 will lead to an approximation, at least, under general specifications for the regression model and the measurement error model. The key concept is to expand the naive likelihood about $z=x$ so that the correction is calculated in terms of $E_{Z|X=x}(z-x)^k$. This works well in the additive case especially if one initially recognizes that the correction depends upon finding functions of $\mathbf{z}$ with expectations, $e^{\beta x}$ and $x\,e^{\beta x}$. To illustrate the method, it can be used to verify that

$$E_{Z|X=x}\left[e^{\beta z-\beta^2\sigma^2/2}\right]=e^{\beta x}\ .$$

More importantly it becomes relatively straightforward to derive approximations which could be used for the multiplicative case.

In the additive case or multiplicative case, expand $e^{\beta z}$ in a Taylor's series to get

$$e^{\beta z} = e^{\beta x}\left(\sum_{k=0}^{\infty} \frac{\beta^k (z-x)^k}{k!}\right).$$

An approximation would result by truncating the series. Taking the expectation under the normal measurement error model, $z \sim \phi((z-x)/\sigma)$, where

$$E_{Z|X=x}(z-x)^k = \begin{cases} 0; & k = 1,3,\cdots \\ (2m-1)\cdots 5\cdot 3\cdot \sigma^{2m}; & k = 2m \end{cases}$$

gives

$$\begin{aligned} E_{Z|X=x}\left[e^{\beta z}\right] &= e^{\beta x}\left(\sum_{k=0}^{\infty} \frac{\beta^k E_{Z|X=x}(z-x)^k}{k!}\right) \\ &= e^{\beta x}\left(\sum_{m=0}^{\infty} \frac{\beta^{2m}(2m-1)\cdots 5\cdot 3\cdot \sigma^{2m}}{(2m)!}\right); m = \frac{k}{2} \\ &= e^{\beta x}\left(\sum_{m=0}^{\infty} \frac{\beta^{2m}\sigma^{2m}}{2^m m!}\right) \\ &= e^{\beta x}\, e^{\beta^2\sigma^2/2} = e^{\beta x + \beta^2\sigma^2/2}. \end{aligned}$$

Note that truncating at $m = 1$ in a multiplicative model leads to the approximation,

$$e^{\beta x} \approx E_{Z|X=x}\left(e^{\beta z}\Big/\left(1 + \tfrac{v^2}{v^2+1} z^2\right)\right).$$

To verify that

$$E_{Z|X=x}\left[z\, e^{\beta z}\right] = x\, e^{\beta x + \beta^2\sigma^2/2},$$

multiply both sides of the previous expansion by z to get

$$z\, e^{\beta z} = e^{\beta x}\left(\sum_{k=0}^{\infty} \frac{\beta^k z(z-x)^k}{k!}\right).$$

For k even,

$$0 = \mathrm{E}_{Z|X=x}(z-x)^{k+1} = \mathrm{E}_{Z|X=x}\left[z(z-x)^{k}\right] - x\mathrm{E}_{Z|X=x}(z-x)^{k}$$

which implies that

$$\mathrm{E}_{Z|X=x}\left[z(z-x)^{k}\right] = x\mathrm{E}_{Z|X=x}(z-x)^{k} \ .$$

Then, by the previous argument,

$$\mathrm{E}_{Z|X=x}\left[z\,\mathrm{e}^{\beta z}\right] = x\,\mathrm{e}^{\beta x}\left(\sum_{k=0}^{\infty}\frac{\beta^{k}\,\mathrm{E}_{Z|X=x}(z-x)^{k}}{k!}\right) = x\,\mathrm{e}^{\beta x+\beta^{2}\sigma^{2}/2} \ .$$

A1-3 Expression for the Approximation of a Replacement Method to the Structural Likelihood.

In this section, the notation will be condensed to simplify expressions and more importantly to make the arguments uncluttered by unnecessary notation. Consider the structural likelihood associated with (y,z) where $\mu = g(\eta)$, $\eta = \alpha + \beta x$ and the other distributions satisfy the definitions of Section 2.1. Then this structural likelihood will be denoted as

$$L(\mu; y, z) = E_{X|Z=z}\left[\overset{*}{L}(\mu; y, x)\right].$$

Let the log-likelihood be denoted by $\ell(\cdot)$ and let the expected value, $\xi_G(z)$, be denoted simply as ξ, so that

$$L(\mu; y, z) = E_{X|Z=z}\left[\exp\left(\overset{*}{\ell}(\mu; y, x)\right)\right]$$

$$= E_{X|Z=z}\left[\exp\left(\overset{*}{\ell}(\mu; y, x) - \overset{*}{\ell}(\mu; y, \xi) + \overset{*}{\ell}(\mu; y, \xi)\right)\right].$$

A Taylor's expansion of $\overset{*}{\ell}(\mu; y, x)$ about $x = \xi$ gives

$$\overset{*}{\ell}(\mu; y, x) - \overset{*}{\ell}(\mu; y, \xi) \approx \left(\beta\, g'(\eta)\frac{\partial \overset{*}{\ell}(\mu; y, x)}{\partial \mu}\right)\Bigg|_{x=\xi} (x - \xi)$$

and substitution into the previous expression yields

$$\ell(\mu; y, z) - \overset{*}{\ell}(\mu; y, \xi) \approx \ln E_{X|Z=z}\left[\exp\left(\beta\, g'(\eta)\frac{\partial \overset{*}{\ell}(\mu; y, \xi)}{\partial \mu}(x - \xi)\right)\right]$$

$$= \ln \mathrm{E}_{X|Z=z}\left[\sum_{k=0}^{\infty}\left(\beta\, g'(\eta)\,\frac{\partial \overset{*}{\ell}(\mu;y,\xi)}{\partial\mu}(x-\xi)\right)^{k}\Big/ k!\right]$$

$$\approx \ln \mathrm{E}_{X|Z=z}\left[1+\beta\, g'(\eta)\,\frac{\partial \overset{*}{\ell}(\mu;y,\xi)}{\partial\mu}(x-\xi)+\tfrac{1}{2}\left(\beta\, g'(\eta)\,\frac{\partial \overset{*}{\ell}(\mu;y,\xi)}{\partial\mu}(x-\xi)\right)^{2}\right]$$

$$\approx \ln\left[1+\tfrac{1}{2}\left(\beta\, g'(\eta)\,\frac{\partial \overset{*}{\ell}(\mu;y,\xi)}{\partial\mu}\right)^{2}\mathrm{E}_{X|Z=z}(x-\xi)^{2}\right].$$

In the absence of measurement errors, $\ell(\mu;y,z)-\overset{*}{\ell}(\mu;y,\xi)=0$ so continuity implies that the difference in these log-likelihoods will be small for sufficiently small measurement errors and the above expression quantifies that error whenever the measurement errors are small. In applications, the expected difference is of interest; i.e. $\mathrm{E}_{Y}\left[\ell(\mu;y,z)-\overset{*}{\ell}(\mu;y,\xi)\right]$. Noting that $\ln(1+\delta)\approx\delta$ for small δ, $\mathrm{E}_{X|Z=z}(x-\xi)^{2}=V_{\mathrm{G}}(z)$ and

$$\mathrm{E}_{Y}\left(\frac{\partial \overset{*}{\ell}(\mu;y,\xi)}{\partial\mu}\right)^{2}=\frac{1}{\mathrm{var}(y)},$$

gives

$$\mathrm{E}_{Y}\left[\ell(\mu;y,z)-\overset{*}{\ell}(\mu;y,\xi)\right]\approx\frac{\beta^{2}\left[g'(\eta)\right]^{2}V_{\mathrm{G}}(z)}{2\,\mathrm{var}(y)}.$$

A1-4 Algebraic Details of Section 4.3

Let $f_{Z|X}(\cdot)$ be a lognormal distribution such that

$$E_{Z|X=x}(z) = x \quad \text{and} \quad V_{Z|X=x}(z) = (\nu x)^2 ,$$

then $\ln(z)$ has a normal distribution with mean,

$$\mu = \ln(x) - \ln\left(\sqrt{\nu^2 + 1}\right) ,$$

and variance,

$$\sigma^2 = \ln\left(\nu^2 + 1\right) .$$

Proof: If $\ln(z)$ has a normal distribution with mean, μ, and variance, σ^2, then z has a lognormal distribution with the moments,

$$E\left(z^k\right) = \exp\left(k\mu + \tfrac{1}{2}k^2\sigma^2\right).$$

Since the converse holds, this gives

$$E_{Z|X=x}(z) = \exp\left(\left(\ln(x) - \ln\left(\sqrt{\nu^2 + 1}\right)\right) + \tfrac{1}{2}\ln\left(\nu^2 + 1\right)\right)$$

$$= x\frac{\sqrt{\nu^2 + 1}}{\sqrt{\nu^2 + 1}} = x$$

and

$$E_{Z|X=x}\left(z^2\right) = \exp\left(2\left(\ln(x) - \ln\left(\sqrt{\nu^2 + 1}\right)\right) + \tfrac{2^2}{2}\ln\left(\nu^2 + 1\right)\right)$$

$$= x^2\frac{\left(\nu^2 + 1\right)^2}{\nu^2 + 1} = x^2\left(\nu^2 + 1\right)$$

which gives

$$V_{Z|X=x}(z) = E_{Z|X=x}(z^2) - \left[E_{Z|X=x}(z)\right]^2 = x^2(\nu^2+1) - x^2 = (\nu x)^2 .$$

- In addition, let $G(\cdot)$ be a mixture of lognormal distributions, $G_k(\cdot)$, with

$$G(x) = pG_1(x) + (1-p)G_2(x) ,$$

where $G_k(\cdot)$ has the parameters, θ_k and ω_k^2, then the following results can be demonstrated:

1. $f_Z(z;G_k)$ is a lognormal distribution with parameters, $\theta_k - \frac{1}{2}\sigma^2$ and $\sigma^2 + \omega_k^2$, where $\sigma^2 \equiv \ln(\nu^2+1)$;

2. $f_{X|Z=z}(x;G_k)$ is a lognormal distribution with parameters, $\theta(z)$ and κ^2, where

$$\theta(z) = \frac{\sigma^2\theta_k + \omega_k^2\left(\ln(z) + \frac{1}{2}\sigma^2\right)}{\sigma^2 + \omega_k^2}$$

and

$$\kappa^2 = \frac{\sigma^2\omega_k^2}{\sigma^2 + \omega_k^2} ;$$

3. $\xi_{G_k}(z) = \exp\left[\theta(z) + \kappa^2/2\right]$;

4. $V_{G_k}(z) = \exp\left[2\theta(z) + 2\kappa^2\right] - \left[\xi_{G_k}(z)\right]^2$;

5. $f_Z(z;G) = pf_Z(z;G_1) + (1-p)f_Z(z;G_2)$;

6. $\xi_G(z)$ and $V_G(z)$ can be written as

$$f_Z(z;G)\,\xi_G(z) = p\,f_Z(z;G_1)\,\xi_{G_1}(z) + (1-p)\,f_Z(z;G_2)\,\xi_{G_2}(z) ,$$

$$f_Z(z;G)V_G(z) = p\, f_Z(z;G_1)V_{G_1}(z) + (1-p)\, f_Z(z;G_2)V_{G_2}(z)\ .$$

Proof of 1:

By definition,

$$f_Z(z;G_k) = \int f_{Z|X}(z|x)\,dG_k(x)$$

$$= \int \frac{1}{\sqrt{2\pi}\,z}\exp\left[-\frac{\left(\ln(z)-\left(\ln(x)-\frac{1}{2}\sigma^2\right)\right)^2}{2\sigma^2}\right]\ \frac{1}{\sqrt{2\pi}\,x}\exp\left[-\frac{\left(\ln(x)-\theta_k\right)^2}{2\omega_k^2}\right]dx\ .$$

Letting $u = \ln(x)$ and $v = \ln(z)+\frac{1}{2}\sigma^2$, transform to normal distributions; that is,

$$f_Z\left(\exp\left(v-\tfrac{1}{2}\sigma^2\right);G_k\right)dv = \frac{dv}{2\pi\sigma\omega_k}\int \exp\left(-\frac{(v-u)^2}{2\sigma^2}\right)\exp\left(-\frac{(u-\theta_k)^2}{2\omega_k^2}\right)du$$

$$= \frac{dv}{2\pi\sigma\omega_k}\int \exp\left(-\frac{1}{2}\left\{\frac{(v-u)^2}{\sigma^2}+\frac{(u-\theta_k)^2}{\omega_k^2}\right\}\right)du\ .$$

It can be shown that

$$\frac{(v-u)^2}{\sigma^2}+\frac{(u-\theta_k)^2}{\omega_k^2} = \frac{\left(u-\theta(z)\right)^2}{\kappa^2}+\frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2}$$

and

$$\sigma^2\omega_k^2 = \kappa^2\left(\sigma^2+\omega_k^2\right)$$

where

$$\theta(z) = \frac{\sigma^2\theta_k + \omega_k^2 v}{\sigma^2 + \omega_k^2} \text{ and } \kappa^2 = \frac{\sigma^2\omega_k^2}{\sigma^2 + \omega_k^2}.$$

Then

$$f_Z(z;G_k)\,dz = \frac{dv}{2\pi\sigma\omega_k}\int \exp\left(-\frac{1}{2}\left\{\frac{(v-u)^2}{\sigma^2} + \frac{(u-\theta_k)^2}{\omega_k^2}\right\}\right)du$$

$$= \frac{dv}{2\pi\sqrt{\kappa^2(\sigma^2+\omega_k^2)}}\int \exp\left(-\frac{1}{2}\left\{\frac{(u-\theta(z))^2}{\kappa^2} + \frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2}\right\}\right)du$$

$$= \frac{dv}{\sqrt{2\pi}\sqrt{(\sigma^2+\omega_k^2)}}\exp\left(-\frac{1}{2}\left\{\frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2}\right\}\right)\int \frac{1}{\sqrt{2\pi}\,\kappa}\exp\left(-\frac{1}{2}\left\{\frac{(u-\theta(z))^2}{\kappa^2}\right\}\right)du$$

$$= \frac{dv}{\sqrt{2\pi}\sqrt{(\sigma^2+\omega_k^2)}}\exp\left(-\frac{1}{2}\left\{\frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2}\right\}\right).$$

Transforming back yields the desired result,

$$f_Z(z;G_k) = \frac{1}{\sqrt{2\pi}\sqrt{(\sigma^2+\omega_k^2)z}}\exp\left(-\frac{1}{2}\left\{\frac{(\ln(z)+\frac{1}{2}\sigma^2-\theta_k)^2}{\sigma^2+\omega_k^2}\right\}\right).$$

That is, $f_Z(z;G_k)$ is a lognormal distribution with parameters, $\theta_k - \frac{1}{2}\sigma^2$ and $\sigma^2+\omega_k^2$.

To verify that

$$\frac{(v-u)^2}{\sigma^2} + \frac{(u-\theta_k)^2}{\omega_k^2} = \frac{(u-\theta(z))^2}{\kappa^2} + \frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2},$$

square the terms to obtain a quadratic expression in u,

$$\frac{(v-u)^2}{\sigma^2}+\frac{(u-\theta_k)^2}{\omega_k^2}=\frac{\left(v^2-2uv+u^2\right)}{\sigma^2}+\frac{\left(u^2-2u\theta_k+\theta_k^2\right)}{\omega_k^2}$$

$$=\frac{u^2-2\kappa^2\left(\frac{v}{\sigma^2}+\frac{\theta_k}{\omega_k^2}\right)u}{\kappa^2}+\frac{v^2}{\sigma^2}+\frac{\theta_k^2}{\omega_k^2}\,.$$

Letting

$$\theta(z)=\kappa^2\left(\frac{v}{\sigma^2}+\frac{\theta_k}{\omega_k^2}\right)=\frac{\omega_k^2 v+\sigma^2\theta_k}{\sigma^2+\omega_k^2},$$

gives

$$\frac{u^2-2\kappa^2\left(\frac{v}{\sigma^2}+\frac{\theta_k}{\omega_k^2}\right)u}{\kappa^2}+\frac{v^2}{\sigma^2}+\frac{\theta_k^2}{\omega_k^2}$$

$$=\frac{u^2-2\theta(z)u+\left(\theta(z)\right)^2}{\kappa^2}-\frac{\left(\theta(z)\right)^2}{\kappa^2}+\frac{v^2}{\sigma^2}+\frac{\theta_k^2}{\omega_k^2}$$

$$=\frac{\left(u-\theta(z)\right)^2}{\kappa^2}-\left(\frac{v}{\sigma^2}+\frac{\theta_k}{\omega_k^2}\right)^2\frac{\sigma^2\omega_k^2}{\sigma^2+\omega_k^2}+\frac{v^2}{\sigma^2}+\frac{\theta_k^2}{\omega_k^2}$$

$$=\frac{\left(u-\theta(z)\right)^2}{\kappa^2}-\left(\frac{\omega_k^2 v}{\sigma^2}+2uv+\frac{\sigma^2\theta_k}{\omega_k^2}\right)\frac{1}{\sigma^2+\omega_k^2}+\frac{v^2}{\sigma^2}\left(\frac{\sigma^2+\omega_k^2}{\sigma^2+\omega_k^2}\right)+\frac{\theta_k^2}{\omega_k^2}\left(\frac{\sigma^2+\omega_k^2}{\sigma^2+\omega_k^2}\right)$$

$$=\frac{\left(u-\theta(z)\right)^2}{\kappa^2}+\frac{(v-\theta_k)^2}{\sigma^2+\omega_k^2}.$$

Proof of 2:

From the definition,

$$f_{X|Z=z}(x;G_k)\,dx = \frac{f_{Z|X}(z|x)\,dG_k(x)}{f_Z(z;G_k)}$$

and the factorization developed for the previous result where

$$\int f_{Z|X}(z|x)\,dG_k(x) = f_Z(z;G_k) \int \frac{1}{\sqrt{2\pi}\,\kappa}\exp\left(-\frac{1}{2}\left\{\frac{(u-\theta(z))^2}{\kappa^2}\right\}\right)du,$$

it follows directly that $f_{X|Z=z}(x;G_k)$ is a lognormal distribution with parameters, $\theta(z)$ and κ^2.

Proof of 3 and 4:

For a lognormal distribution with parameters, $\theta(z)$ and κ^2, the mean is

$$E(x) = \exp\left[\theta(z)+\kappa^2/2\right] = \xi_{G_k}(z)\ ;$$

and the variance is

$$E(x^2)-(E(x))^2 = \exp\left[2\theta(z)+2\kappa^2\right]-\left[\xi_{G_k}(z)\right]^2 = V_{G_k}(z)\ .$$

Proof of 5:

From the definition, we obtain

$$f_Z(z;G) = \int f_{Z|X}(z|x)\,dG(x)$$

$$= \int f_{Z|X}(z|x)\left[pG_1(x)+(1-p)G_2(x)\right]$$

$$p\,f_Z(z;G_1)+(1-p)f_Z(z;G_2)\ .$$

Proof of 6:

Consider evaluating the expectation for an arbitrary function, $h(x)$, then

$$E_{X|Z=z}(h(x))=\int h(x)f_{X|Z=z}(x;G)\,dx$$

$$=\frac{\int h(x)f_{Z|X}(z|x)\,dG(x)}{f_Z(z;G)}\ .$$

Then

$$f_Z(z;G)E_{X|Z=z}(h(x))=\int h(x)f_{Z|X}(z|x)\,dG(x)$$

$$=\int h(x)f_{Z|X}(z|x)\left[pG_1(x)+(1-p)G_2(x)\right]$$

$$=p\int h(x)f_{Z|X}(z|x)\,dG_1(x)+(1-p)\int h(x)f_{Z|X}(z|x)\,dG_2(x)$$

$$=p\,f_Z(z;G_1)\frac{\int h(x)f_{Z|X}(z|x)\,dG_1(x)}{f_Z(z;G_1)}+(1-p)f_Z(z;G_2)\frac{\int h(x)f_{Z|X}(z|x)\,dG_2(x)}{f_Z(z;G_2)}$$

$$=p\,f_Z(z;G_1)\int h(x)f_{X|Z=z}(x;G_1)\,dx+(1-p)f_Z(z;G_2)\int h(x)f_{X|Z=z}(x;G_2)\,dx\ .$$

Substitution of $h(x)=x$ and $h(x)=(x-E(x))^2$ give the required formulae:

$$f_Z(z;G)\,\xi_G(z)=p\,f_Z(z;G_1)\,\xi_{G_1}(z)+(1-p)\,f_Z(z;G_2)\,\xi_{G_2}(z)\ ,$$

$$f_Z(z;G)V_G(z)=p\,f_Z(z;G_1)V_{G_1}(z)+(1-p)\,f_Z(z;G_2)V_{G_2}(z)\ .$$

Appendix 2

This Appendix contains a selection of the computer programs that make up the simulations. Because many programs reflect trivial changes to the reported programs, a listing of every program has not been reproduced. There are three types of programs;

1. Programs to generate the data $(\mathbf{x}, \mathbf{y}, \mathbf{z})$, such as XYZ.GSS
2. Programs to estimate (α, β), such as MLE.GSS
3. Programs to compare the estimates, such as ALP_BETA.GSS

With this organization of tasks it is relatively straightforward to investigate new methods or compare methods in different ways. Many of the estimation programs can be adapted to different distributional assumptions without too much reprogramming. Unfortunately this breaks many programs into several subroutines and this may make them a little hard to follow. The Gauss version that I am running requires the procedures that are called to precede the procedure which calls them. Consequently most estimation programs are organized in the order:

1. Procedures that implement the specific measurement error model / parameterization.
2. Procedures that implement the estimation algorithm, such as Newton-Raphson algorithms with step halving.
3. Main procedure that reads in the $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ data, calls the estimation algorithm, and writes the estimates to a data set.

Chapter 4 discusses the simulation methods in detail and introduces the algorithms that are contained in this appendix.

A2-1.1 XYZ.GSS

Gauss program that generates the simulation data sets. The program calls the Gauss procedure, PIF, which is given in Section A2-1.1. This also requires a table, QTABLE.DAT, of Poisson probabilities which was created by the fortran program, QTABLE.FOR. This program is given in Section A2-1.2.

```
/*
The choice of G = p G1 + q G2 is two log-normal dist'ns.  The errors,
f(z|x),
are also log-normal (cv = .5).  E(y) = .5 + .2

    cv = 0.5 ; p = 0.5 ;
    G1:  u1 = LN(0.1) ; sig1 = 1.00 ;
    G2:  u2 = LN(1.0) ; sig2 = 1.00 ;

    This GAUSS program simulates Y from a Poisson(X*BETA) and Z from a
log_normal with constant cv and E(Z) = X.  The input parameters are BETA,
cv,
and k where the sample size is 2^(k-1)*100.

*/

LOADP PATH = E:\ PIF ;
LOAD QP[100,2] = E:\QPTABLE.DAT ;

    CREATE FXYZ = E:\XYZDAT.GDT WITH Y,3,2 ;


/*   Set Simulation Parameters   */

    k = 1 ;

    cv = 0.3 ; p = 0.5 ;
    u1 = LN(0.1) ; sig1 = 1.00 ;
    u2 = LN(3.0) ; sig2 = 0.50 ;
    LET BETA = 0.5 0.2 ;

/*                                */

       n = 2^(k-1) * 100 ;
       SIGMA2 = LN(cv*cv + 1) ;

       Y = ZEROS(N,3) ;

       ITER = 1;
    mm = INT(p * n) ;
    mn = n - mm ;

DO WHILE ITER <= 500 ;
    X = EXP(sig1*RNDN(mm,1) + u1) | EXP(sig2*RNDN(mn,1) + u2) ;
    Y[.,1] = int(100*X) ;
```

```
    M = (ONES(N,1)~X)*BETA ;
    I = 1;
    DO UNTIL I > N ;
    Y[I,2] = PIF(M[I,1],QP);
    I = I + 1;
    ENDO;

    Y[.,3] = Y[.,1] .* EXP(SQRT(SIGMA2)*RNDN(n,1) - SIGMA2/2) ;


    CHK = WRITER(FXYZ,Y) ;
    IF CHK /= N ;
       ERRORLOG "PROBLEM WRITING Y" ;
       FXYZ = CLOSE(FXYZ) ;
       ENDIF ;

ITER ;
ITER=ITER+1 ;
ENDO;
    FXYZ = CLOSE(FXYZ) ;

END;
```

A2-1.2 PIF.GSS

Gauss program that implements the PIF algorithm to generate Poisson samples.

```
PROC PIF(Mu,QP) ;

  /*
This is the PIF algorithm (Fishman, 1976)to generate random Poisson
samples with mean, Mu, when Mu <= 100.  When Mu > 100 it returns

        MAX(0,INT(Mu + z*SQRT(Mu) +.5))

where z is a random normal (0,1) deviate.

QP[100,2] is a table with Poisson probabilities:

          QP[i,1] = Prob[X <= i | Mu = i]
          QP[i,2] = Prob[X = i | Mu = i]

This table is created by a Fortran program, QPTABLE.FOR


Reference:

   Fishman, G S.  1976.  Sampling from the Poisson Distribution on a
Computer.  Computing 17:147-156.

  */
LOCAL M,D,E,U,X,F,V,A,B ;
IF Mu <= 100;
M = FLOOR(Mu) ;
F = Mu - M ;
X = M ;
   IF M > 0 ;
   D = QP[M,1] ;
   E = QP[M,2] ;
   U = RNDU(1,1) ;
      IF U <= D ;
      D = D - E ;
          DO UNTIL ((U > D) + (X .EQ 0)) ;
          E = E*X/M ;
          X = X - 1 ;
          D = D - E ;
          ENDO ;
      ELSE ;
          DO UNTIL U <= D ;
          X = X + 1;
          E = E*M/X ;
          D = D + E ;
          ENDO;
      ENDIF ;
```

```
        ENDIF ;
        V = 0 ;
           IF F > 0 ;
           A = EXP(- F) ;
           B = A ;
           U = RNDU(1,1) ;
               DO UNTIL U <= A ;
               V = V + 1 ;
               B = B*F/V ;
               A = A + B ;
               ENDO ;
           ENDIF ;
        X = V + X ;
    ELSE ;
    X = MAXC(0 | INT(Mu+RNDN(1,1)*SQRT(Mu)+.5)) ;
    ENDIF;
    RETP(X) ;
    ENDP;

    LOAD QP[100,2] = I:QPTABLE.DAT ;
    I = 1;
    Y = ZEROS(25,2) ;
    DO UNTIL I > 25 ;
    M = 10*I+RNDU(1,1) ;
    PSN = PIF(M,QP);
    Y[I,.] = PSN ~ M ;
    I = I + 1;
    ENDO;
Y ;
MEANC(Y) ; STDC(Y) ;
END;
```

A2-1.3 QTABLE.FOR

Generates the table of Poisson Probabilities required by PIF.GSS.

```
      DOUBLE PRECISION A,B
      OPEN(UNIT=6,FILE='QTABLE.DAT')
      WRITE(*,5)
    5 FORMAT(' N? ')
      READ(*,*) N
      DO 20 I = 1,N
      U = I
      A = EXP(-U)
      B = A
      DO 10 K = 1,I
      X = K
      B = B*U/X
   10 A = A + B
   20 WRITE(6,30) A,B
   30 FORMAT(' ',2(2X,F10.8))
      STOP
      END
```

A2-2.1 MLE.GSS

Gauss program that calculates the maximum likelihood estimates under the quasi-structural model.

```
NEW;
DISABLE ;
cv = .3 ; n = 1600 ;
LET BETA = .5 .2 ;

    OUTPUT FILE = E:\V30K5B02\MLE.LST RESET ;
      OPEN FXYZ = E:\V30K5B02\XYZDAT.GDT ;


DECLARE PSNERR[1909] ?= 0 ;
LOADEXE PSNERR = E:\PSNERR.EXE ;
FORMAT /RD 8,3 ;


    PROC (3) = CALCLIKE(b,X,Y,Z,cv,n) ;
    /*
    This procedure is called by PROC MLE_ERR.  It calculates
    the deviance, score, and information in a Poisson(a+bx)
    model with lognormal f(z|x).
    */
        LOCAL DEV,SCORE,INFORM ;
        DEV = 0 ; SCORE = ZEROS(2,1) ; INFORM = ZEROS(2,2) ;
        CALLEXE PSNERR(b,X,Y,Z,n,cv,DEV,SCORE,INFORM) ;
        RETP(DEV,SCORE,INFORM) ;
        ENDP ;


    PROC (3) = MLE_NR(X,Y,Z,cv,START) ;
    LOCAL QUIT,HALF,b,OLD_DEV,DEV,SCORE,INFO,CHANGE,n,STEP,V,NS ;
    QUIT = .000001 ;
    HALF = 10 ;

    /* This is the Newton-Raphson algorithm with step halving.
       QUIT - sets the maximum change in the deviance that
       is considered convergence.
       HALF - sets the maximum number of step halvings attempted
       before returning.

       This program returns the mle's, b, their covariance matrix, V,
       and a flag for convergence,
          0  => did not converge
          DEV  => converged

       This program needs two other PROC's to define the likelihoods
       and starting values.

    */
```

```
    b = START ;
    n = ROWS(Y) ;
    {OLD_DEV,SCORE,INFO} = CALCLIKE(b,X,Y,Z,cv,n) ;
    CHANGE = 1 + QUIT ;
    DO UNTIL CHANGE < QUIT ;
    V = INV(INFO) ;
    STEP = V*SCORE ;
    b = b + STEP ;
    NS = 0 ;
    LIKE: {DEV,SCORE,INFO} = CALCLIKE(b,X,Y,Z,cv,n) ;
    IF DEV > OLD_DEV ;
       STEP = STEP/2 ;
       b = b - STEP ;
       NS = NS + 1 ;
       IF NS > HALF ;
             RETP(b - STEP,V,0) ;
             ENDIF ;
       GOTO LIKE ;
       ENDIF ;
    CHANGE = OLD_DEV - DEV ;
    OLD_DEV = DEV ;
    ENDO ;
    RETP(b,INV(INFO),DEV) ;
    ENDP ;


    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       ITER = ITER + 1 ;
       IN = READR(FXYZ,n) ;

       X = IN[.,1] / 100 ;
       X = X + (X .== 0) * .005 ;
       Y = IN[.,2];
       Z = IN[.,3] / 100 ;
       Z = Z + (Z .== 0) * .005 ;


       {B,VB,CONVERGE} = MLE_NR(X,Y,Z,cv,BETA) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUTPUT ON ;
       OUT ;
       OUTPUT OFF ;


       ENDO ;

    FXYZ = CLOSE(FXYZ) ;

END;
```

A2-2.2 PSNERR.FOR

```
      SUBROUTINE PSNERR(BETA,XX,YY,ZZ,R,CV,DEV,S,V)
      REAL*8 XX(1),C1,C2,CV,F,YF
      REAL*8 BETA(1),YY(1),ZZ(1),R,DEV,S(1),V(2,2),X2
      REAL*8 A,B,C,AVE,W,TEMP,SW,SC,D(2),D2(2,2),DS(2,2),X,Y,YI,Z
      INTEGER*2 I,J,K,N
      N = R
      A = BETA(1)
      B = BETA(2)
      C1 = SQRT(CV*CV+1)
      C2 = 2*DLOG(CV*CV+1)
C
C  This assumes:
C     - f(z|x) is lognormal with E(z|x) = x & constant cv.
C     - DEV, S, & V are initially zero
C     - X & Z are greater than 0
C     - Y greater than or equal to 0
C
C   Returns:
C     - deviance as DEV
C     - score as S
C     - information as V
C
      DO 40 I = 1,N
C  Set i values
      Y = YY(I)
      Z = ZZ(I)
      SW = 0.0
      SC = 0.0
      YI = 0.0
      K = Y
      F = 0.0
      IF(K.LE.1) GO TO 16
      DO 15 J=2,K
      YF = J
   15 F = F + DLOG(YF)
   16 CONTINUE
      IF(Y.GT.0) YI = Y*DLOG(Y) - Y - F
      DO 20 J = 1,2
      D(J) = 0.0
      DO 20 K = 1,2
      DS(J,K) = 0.0
   20 D2(J,K) = 0.0
C
C  Sum over j=1 to n
C
      DO 30 J = 1,N
      X = XX(J)
      X2 = X*X
      C = DLOG(C1*Z/X)
      C = DEXP(-C*C/C2)
```

```
      SC = SC + C
      AVE = A+B*X
      W = C*DEXP(Y*DLOG(AVE)-AVE-F)
      SW = SW + W
      TEMP = Y/AVE - 1
      D(1) = D(1) + TEMP*W
      D(2) = D(2) + TEMP*W*X
      TEMP = TEMP*TEMP*W
      D2(1,1) = D2(1,1) + TEMP
      D2(1,2) = D2(1,2) + TEMP*X
      D2(2,2) = D2(2,2) + TEMP*X2
      TEMP =  W * Y/(AVE*AVE)
      DS(1,1) = DS(1,1) + TEMP
      DS(1,2) = DS(1,2) + TEMP*X
   30 DS(2,2) = DS(2,2) + TEMP*X2
      D2(2,1) = D2(1,2)
      DS(2,1) = DS(1,2)
      DEV = DEV + YI - DLOG(SW/SC)
      DO 40 J = 1,2
      S(J) = S(J) + D(J)/SW
      DO 40 K = 1,2
   40 V(J,K) = V(J,K) - D2(J,K)/SW + DS(J,K)/SW + (D(J)/SW)*(D(K)/SW)
      DEV = 2*DEV
      RETURN
      END
```

A2-2.3 WLS_PSN.GSS

The iteratively reweighted least squares estimates and replacement estimates were calculated using this algorithm, i.e. structural, quasi-structural, and modified Whittemore. By setting V = 0, this algorithm is also utilized to calculate Schafer's estimates, the naive estimates, as well as the Poisson regression estimates based upon *x* known.

```
PROC(3) = WLS_PSN(beta,Y,E,V) ;

     local n,b,diff,i,W,XW,X ;

     n = ROWS(Y) ;
     b = beta ;
     i = 0 ;
     diff = b ;
     X = ONES(n,1) ~ E ;

     do while ABS(SUMC(diff)) > 0.0001 ;
     diff = b ;
     i = i + 1 ;
        if i > 20 ;
        RETP(b,INV(XW'X),-1) ;
        endif ;
     W = X*b + b[2,1]*b[2,1]*V ;
     XW = (ONES(n,1) ./ W) ~ (E ./ W) ;
     b = INV(XW'X)*XW'Y ;
     diff = diff - b ;
     endo ;

     RETP(b,INV(XW'X),i) ;
    ENDP ;
```

A2-2.4 EVX_Z.FOR

```
      SUBROUTINE EVX_Z(E,V,ZZ,XX,CV,RM,RN)
      REAL*8 E(1),V(1),ZZ(1),XX(1),RM,RN,CV
      REAL*8 Z,X,U
      INTEGER*2 N,M,I,J
C
C     M = # of nonzero Z's
C     N = # of nonzero X's
C
      M = RM
      N = RN
      SIGMA2 = DLOG(CV*CV+1)
      DO 10 I = 1,M
      Z = DLOG(ZZ(I))
      SX = 0.0
      SX2 = 0.0
      S = 0.0
      DO 20 J = 1,N
      X = XX(J)
      U = Z - DLOG(X) + SIGMA2 / 2.0
      U = DEXP(-.5*U*U/SIGMA2)
      S = S + U
      SX = SX + X*U
   20 SX2 = SX2 + X*X*U
      E(I) = SX / S
   10 V(I) = SX2 / S - E(I)*E(I)
      RETURN
      END
```

A2-2.5 EVX_X.GSS

```
NEW;
DISABLE ;
CV = .3 ; n = 800 ;
LET BETA = .5 .2 ;

      OPEN FXYZ = E:\V30K4B02\XYZDAT.GDT ;

DECLARE EVx_z[1789] ?= 0 ;
LOADEXE EVx_z = E:\EVx_z.EXE ;
LOADP WLS_PSN ;
FORMAT /RD 8,3 ;

    E = ZEROS(n,1) ;
    V = ZEROS(n,1) ;
    VV = ZEROS(n,1) ;
    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       ITER = ITER + 1 ;
       IN = READR(FXYZ,n) ;

       X = IN[.,1] / 100 ;
       X = (X.==0) * 0.005 +  X ;

       Y = IN[.,2];
       Z = IN[.,3] / 100;
       Z = (Z.==0) * 0.005 +  Z ;

       CALLEXE EVx_z(E,V,Z,X,cv,n,n) ;

       OUTPUT FILE = E:\V30K4B02\EVx_z.LST ON ;
       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;
       OUTPUT OFF ;

       OUTPUT FILE = E:\V30K4B02\Ex_z.LST ON ;
       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,VV) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;
       OUTPUT OFF ;

       ENDO ;

    FXYZ = CLOSE(FXYZ) ;

END;
```

A2-3.1 MIXTURE.GSS

```
cv = .3 ; n = 800 ;
LET BETA = .5 .2 ;

    OUTPUT FILE = E:\V30K4B02\MIX.LST RESET ;
      OPEN FXYZ = E:\V30K4B02\XYZDAT.GDT ;

    p = 0.5 ;
    u1 = LN(0.1) ; sig1 = 1.00 ;
    u2 = LN(3.0) ; sig2 = 0.50 ;

LOADP WLS_PSN ;
FORMAT /RD 8,3 ;

    sig = LN(cv*cv+1) ;
    sig1 = sig1*sig1 + sig ;
    sig2 = sig2*sig2 + sig ;

    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       ITER = ITER + 1 ;
       IN = READR(FXYZ,n) ;
       Y = IN[.,2];
       Z = IN[.,3] / 100;
       Z = (Z .== 0) * 0.005 + Z ;

       /* Mixture of Lognormal Distributions */

       Z = LN(Z) + sig/2;
       B = sig / sig1 ;
       f = EXP(-.5*(Z - u1)^2/sig1) / SQRT(sig1) ;
       E = p*f.*EXP((1-B)*(Z + sig/2) + B*u1) ;
       V = p*f.*EXP(2*((1-B)*(Z + sig) + B*u1)) ;
       fG = p*f ;
       B = sig / sig2 ;
       f = EXP(-.5*(Z - u2)^2/sig2) / SQRT(sig2) ;
       fG = fG + (1-p)*f ;
       E = (E + (1-p)*f.*EXP((1-B)*(Z + sig/2) + B*u2)) ./ fG ;
       V = (V + (1-p)*f.*EXP(2*((1-B)*(Z + sig) + B*u2))) ./ fG ;

       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;

       ENDO ;

    FXYZ = CLOSE(FXYZ) ;

END;
```

A2-3.2 LN_NORM.GSS

```
cv = .3       ; n = 1600 ;
LET BETA = .5 .1 ;

    OUTPUT FILE = E:\V30K5B01\LN_NORM.LST RESET ;
      OPEN FXYZ = E:\V30K5B01\XYZDAT.GDT ;

LOADP PATH=C:\GAUSS\CP WLS_PSN ;
FORMAT /RD 8,3 ;

    sig2 = LN(cv*cv+1) ;
    theta = 0.151 * sig2 ;
    w = 0.924 ;
    w = w*w ;
    k = sig2*w/(sig2 + w) ;

    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       ITER = ITER + 1 ;
       IN = READR(FXYZ,n) ;
       Y = IN[.,2];
       Z = IN[.,3] / 100;
       Z = (Z .== 0) * 0.005 + Z ;

       /* Structural Model with multiplicative errors */
        E = (theta + w * (ln(Z) + 0.5*sig2))/(sig2 + w) ;
        V = exp(2*(E+k)) ;
        E = exp(E + 0.5*k) ;
        V = V - E.*E ;

       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;

       ENDO ;

    FXYZ = CLOSE(FXYZ) ;

END;
```

A2-4.1 ATTN.GSS

```
  OUTPUT FILE = E:\V30K4B02\ATTN.LST RESET ;
    OPEN FXYZ = E:\V30K4B02\XYZDAT.GDT ;
n = 800 ;
cv = 0.3 ;
LET BETA = .5 .2 ;

/*  No changes past here */
LOADP WLS_PSN ;
FORMAT /RD 8,3 ;
    c = cv*cv/(cv*cv+1) ;
    E = ZEROS(n,1) ;
    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       IN = READR(FXYZ,n) ;
       ITER = ITER + 1 ;
       Z = IN[.,3] / 100 ;
       Z = (Z .== 0) * .005 + Z ;
       Zbar = SUMC(Z) / n ;
       Y = IN[.,2];

    {B,V,CONVERGE} = WLS_PSN(BETA,Y,Z,E) ;
i=0;
b0=b[2,1];
do while i<5;
i=i+1;
    W = 1 / (B[1,1] + B[2,1]*Z) ;
    SS = W'(Z-Zbar)^2 ;
    SS = SS/(SS - c*W'Z^2) ;
    B[2,1] = B0 * SS ;
    B[1,1] = (W'Y - B[2,1]*W'Z)/SUMC(W) ;
endo;
    OUT = B ~ SQRT(DIAG(V)) ~ V ~ (ITER|CONVERGE) ;
    OUT ;
       ENDO ;

    FXYZ = CLOSE(FXYZ) ;
END;
```

A2-5.1 NAK2MURA.GSS

```
NEW;
DISABLE ;
CV = .3 ; n = 800 ;
LET BETA = .5 .2 ;

    OUTPUT FILE = E:\V30K4B02\NAKA2.LST RESET ;
      OPEN FXYZ = E:\V30K4B02\XYZDAT.GDT ;

    PROC (3) = NAK2LIKE(BETA,Y,Z,CV,n) ;
       local a,b,v,c2,Ey,den2,den3,den4,s1,s2,score,i1,i2,i3,inform ;
       a = beta[1,1] ;
       b = beta[2,1] ;
       v = cv*cv + 1 ;
       c2 = z^2 * (v - 1)/v ;
       Ey = (ONES(n,1)~z) * beta ;
       den2 = a^2 + 2*a*b*z + b^2*z^2/v ;
        den3 = a^3 + 3*a^2*b*z + 3*a*b^2*z^2/v + b^3*z^3/v^3 ;
       den4 = a^4 + 4*a^3*b*z + 6*a^2*b^2*z^2/v
              + 4*a*b^3*z^3/v^3 + b^4*z^4/v^6 ;

       s1 = SUMC( y ./ Ey - 1 ) - b^2*y'(c2./den3) ;
       s2 = z'( y ./ Ey - 1 ) + a*b*y'(c2./den3) ;
       score = s1 | s2 ;

       i1 = SUMC(y./den2) + 3*b^2*y'(c2./den4) ;
       i2 = z'(y./den2) + b*y'((b*z-2*a).*c2./den4) ;
       i3 = (z^2)'(y./den2) - a*y'((b*z-2*a).*c2./den4) ;
       inform = (i1~i2) | (i2~i3) ;

    RETP(ABS(SUMC(SCORE)),SCORE,INFORM) ;
    ENDP ;


    PROC (3) = NAKAMURA(BETA,Y,Z,CV) ;

    /*
    This algorithm minimizes the Poisson deviance for a two parameter
    model.  The actual parameterization depends upon the Fortran
    subroutine, POISSON, which is called by this PROC.  This is a
    Newton-Raphson algorithm with step halving.

    QUIT - specifies the convergence bound on changes in deviance
    MAX  - specifies the maximum # of iterations
    HALF - specifies the maximum # of step halvings
    */

    LOCAL QUIT,MAX,HALF,NS,ITR,n,V,STEP,CHANGE,b,
      OLD_DEV,DEVIANCE,SCORE,INFORM ;

    QUIT = .00001 ;
```

```
    MAX = 10 ;
    HALF = 5 ;

    n = ROWS(Z) ;
    b = BETA ;

    {OLD_DEV,SCORE,INFORM} = NAK2LIKE(b,Y,Z,CV,n) ;

    DEVIANCE = 0 ; ITR = 0 ; CHANGE = QUIT + 1 ;
    DO WHILE CHANGE > QUIT ;
    V = INV(INFORM) ;
       ITR = ITR + 1 ;
       IF ITR > MAX ;
       RETP(b,V,0) ;
       ENDIF ;

    STEP = V*SCORE ;
    b = b + STEP ;
    {DEVIANCE,SCORE,INFORM} = NAK2LIKE(b,Y,Z,CV,n) ;
    CHANGE = OLD_DEV - DEVIANCE ;

       NS = 0 ;
       DO WHILE CHANGE < 0 ;

      NS = NS + 1 ;
      IF NS > HALF ;
      RETP(b,V,0) ;
      ENDIF ;

       STEP = STEP/2 ;
       b = b - STEP ;
       {DEVIANCE,SCORE,INFORM} = NAK2LIKE(b,Y,Z,CV,n) ;
       CHANGE = OLD_DEV - DEVIANCE ;
       ENDO ;

    OLD_DEV = DEVIANCE ;
    ENDO ;

    RETP(b,V,DEVIANCE) ;
    ENDP ;


FORMAT /RD 8,3 ;
    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       XYZ = READR(FXYZ,n) ;
       ITER = ITER + 1 ;
       Y = XYZ[.,2] ;
       Z = XYZ[.,3]/100 ;
       Z = (Z .== 0) * .005 + Z ;

    {B,V,CONVERGE} = NAKAMURA(BETA,Y,Z,CV) ;
    OUT = B ~ SQRT(DIAG(V)) ~ V ~ (ITER|CONVERGE) ;
    OUT ;
```

```
        ENDO ;

    FXYZ = CLOSE(FXYZ) ;
END;
```

A2-6.1 WLS_SIM.GSS

```
cv = .3 ; n = 800 ;
LET BETA = .5 .1 ;

      OPEN FXYZ = C:\SIMULATN\24JUL92\V30K4B01\XYZDAT.GDT ;

LOADP QSH,WLS_PSN ;
FORMAT /RD 8,3 ;

    sig = LN(cv*cv+1) ;

    ITER = 0 ;
    DO UNTIL EOF(FXYZ) ;
       ITER = ITER + 1 ;
       IN = READR(FXYZ,n) ;
       X = IN[.,1] / 100 ;
       X = (X .== 0) * .005 + X ;
       Y = IN[.,2];
       Z = IN[.,3] / 100 ;
       Z = Z + (Z .== 0) * .005 ;

       V = ZEROS(n,1) ;
       OUTPUT FILE = C:\SIMULATN\24JUL92\V30K4B01\PSN.LST ON ;
       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,X,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;
       OUTPUT OFF ;

       OUTPUT FILE = C:\SIMULATN\24JUL92\V30K4B01\PSNZ.LST ON ;
       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,Z,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;
       OUTPUT OFF ;

       OUTPUT FILE = C:\SIMULATN\24JUL92\V30K4B01\Schafer.LST ON ;
       E = QSH(Z,cv);
       {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,V) ;
       OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
       OUT ;
       OUTPUT OFF ;

       /* Whittemore Method with multiplicative errors */
       Z = LN(Z) ;
       w = MEANC(Z) ;
       v2 = SUMC((Z-w)^2) / n ;
       B = sig2 / v2 ;
       v2 = (1-B) * sig2 ;
       E = (1-B)*(Z + .5*sig2) + B*w ;
       V = EXP(2*(E + v2)) ;
       E = EXP(E + .5*v2) ;
       V = V - E .* E ;
```

```
        OUTPUT FILE = C:\SIMULATN\24JUL92\V30K4B01\Whit.LST ON ;
        {B,VB,CONVERGE} = WLS_PSN(BETA,Y,E,V) ;
        OUT = B ~ SQRT(DIAG(VB)) ~ VB ~ (ITER|CONVERGE) ;
        OUT ;
        OUTPUT OFF ;


        ENDO ;

    FXYZ = CLOSE(FXYZ) ;

END;
```