Measurement Error Regression With Unknown Link: Dimension Reduction and Data Visualization

RAYMOND J. CARROLL and KER-CHAU LI*

A general nonlinear regression problem is considered with measurement error in the predictors. We assume that the response is related to an unknown linear combination of a multidimensional predictor through an *unknown* link function. Instead of observing the predictor, we instead observe a surrogate with the property that its expectation is linearly related to the true predictor with constant variance. We identify an important transformation of the surrogate variable. Using this transformed variable, we show that if one proceeds with the usual analysis ignoring measurement error, then both ordinary least squares and sliced inverse regression yield estimates which consistently estimate the true regression parameter, up to a constant of proportionality. We derive the asymptotic distribution of the estimates. A simulation study is conducted applying sliced inverse regression in this context.

KEY WORDS: Data visualization; Dimension reduction; Errors in variables; Generalized linear model; Logistic regression; Sliced inverse regression.

1. INTRODUCTION

This article explores estimation of regression coefficients in general nonlinear models when the link function is unspecified and the predictors are measured with error. Errors in the regressors can cause severe bias in estimation unless suitable adjustment has been made. For linear models, there are many references and techniques available (see Fuller 1987). But nonlinear measurement error models have received relatively less attention until recently (see Carroll 1989 and Carroll and Stefanski 1990 for recent reviews). This article will address the nonlinear case from the viewpoint of dimension reduction and data visualization as given in Li (1990a,b, 1991) for the regressor-error-free situation. Applications of binary measurement error models were discussed by Carroll, Spiegelman, Lan, Bailey, and Abbott (1984), Carroll and Wand (1991), Pepe and Fleming (1991), Pierce, Stram, Vaeth, and Schaefer (1992), Rosner, Willett, and Spiegelman (1989), and Tosteson, Stefanski, and Schaefer (1989), among others.

With errors in the regressor variables, nonlinear regression becomes much more complicated. For example, the likelihood function generally involves multiple integration, and issues of model sensitivity and robustness are not well understood. Let $\mathbf{Y}$ be the response, $\mathbf{x}$ the true predictor, and $\mathbf{w}$ the observed surrogate. Likelihood analysis requires specifying functional forms for the distributions of $\mathbf{Y}$ given $\mathbf{x}$ and of $\mathbf{x}$ given $\mathbf{w}$. Part of our goal is to reduce the necessity for fully specifying these functional forms.

When the dimension of $\mathbf{w}$ is high, many techniques are likely to break down even when the regressors are free from error. This problem is compounded by the uncertainty in choosing the correct form for the regression function. For example, in binary regression, it is not clear why a logistic or probit model is the sole choice. For a continuous response variable, it is equally questionable to recommend a Box–Cox transformation rather than a generalized linear model. These issues can be addressed by assuming that the conditional distribution of the response variable $\mathbf{Y}$ given the predictor variable $\mathbf{x}$ depends on $\mathbf{x}$ only through a linear combination of $\mathbf{x}$, $\beta'\mathbf{x}$; equivalently, for some completely unknown link function g and a random variable ε independent of $\mathbf{x}$,

$$\mathbf{Y} = g(\alpha + \beta'\mathbf{x}, \varepsilon) \tag{1.1}$$

(see Brillinger 1977, 1983, Li and Duan 1989). Model (1.1) states that $\mathbf{Y}$ and $\mathbf{x}$ are independent given $\beta'\mathbf{x}$. In the error-free case, these authors found that consistent estimation of the direction of the slope vector β up to a constant of proportionality is possible under a key condition on the design distribution; see (2.2) in Section 2.

Another aspect of estimating β proportionally is related to the issue of dimension reduction and data visualization as addressed by Li (1990a,b, 1991). Li (1991) formulated this visualization problem as a dimension-reduction problem. In his model the conditional density of $\mathbf{Y}$ given $\mathbf{x}$ depends on $\mathbf{x}$ only through a small number of projected variables, $\beta_1'\mathbf{x}, \ldots, \beta_K'\mathbf{x}$:

$$\mathbf{Y} = g(\beta_1'\mathbf{x}, \ldots, \beta_K'\mathbf{x}, \varepsilon), \tag{1.2}$$

where β_k's are unknown vectors and g is an arbitrary function. The problem is to find the space, the edr (effective dimension reduction) space, spanned by these β's, *without going through tedious parametric or nonparametric model-fitting processes.* Under (1.2) the projections of $\mathbf{x}$ along vectors in the edr space contain all information about the relationship between $\mathbf{Y}$ and $\mathbf{x}$. When K equals 1, (1.2) reduces to the usual nonlinear regression model (1.1) and the edr space is spanned by the regression slope vector β. Hence estimating β up to a constant of proportionality is sufficient for dimension reduction and data visualization, because a plot of $\mathbf{Y}$ against the estimated variable $\hat{\beta}'\mathbf{x}$ can be quite informative.

In this article, we shall develop techniques for estimating the regression slope up to a constant of proportionality when the regressor is subject to error, *without knowledge of the functional form of g.* In effect, our method provides a simple,

* Raymond J. Carroll is Professor of Statistics, Texas A&M University, College Station, TX 77843-3143 and Visiting Scientist, Epidemiology Methods Section, National Cancer Institute, Bethesda, MD 20892. His research was supported by a grant from the National Cancer Institute. Ker-Chau Li is Professor of Mathematics, UCLA, Los Angeles, CA 90024. His research was supported by NSF Grant DMS89-02494.

easily computed sensitivity analysis to judge the potential effect of misspecifying a model for $\mathbf{Y}$ given $\mathbf{x}$. The basic technical tool is a simple linear transformation $\mathbf{u} = L\mathbf{w}$ of $\mathbf{w}$, where $L = \text{cov}(\mathbf{x}, \mathbf{w})\Sigma_{\mathbf{w}}^{-1}$ and $(\Sigma_{\mathbf{x}}, \Sigma_{\mathbf{w}})$ are the covariance matrices of $(\mathbf{x}, \mathbf{w})$. We show that with this linearly transformed variable $\mathbf{u}$, we may proceed with the analysis as if the errors are free in $\mathbf{u}$. In particular, the usual linear least squares method and sliced inverse regression (SIR) can be applied to $\mathbf{Y}$ against $\mathbf{u}$ to yield a root n consistent estimate under suitable conditions. We can interpret $\mathbf{u}$ as the linear least squares prediction of $\mathbf{x}$ given $\mathbf{w}$. When necessary, the transformation L can be estimated from a validation sample.

In Section 2 we formulate the measurement error problem when the link function is completely unknown. We study the inverse regression curve; that is, the conditional mean of the surrogate variable $\mathbf{w}$ given the response variable $\mathbf{Y}$. Under the design condition (2.2), we show that the inverse regression curve degenerates to a straight line. This property extends to the curve of the conditional mean of $\mathbf{u}$ given $\mathbf{Y}$. Based on the transformed regressor $\mathbf{u}$ and proceeding as if the regressor were error free, we show Fisher consistency for two methods of estimating the direction of β: linear least squares and the SIR estimate. Variants of SIR, like those in Duan and Li (1991), also work, but will not be treated here. Condition (2.2) is further discussed in Remarks 2.2 and 2.3.

In Section 3 we study the large sample properties of our estimates. Asymptotic normality with estimable asymptotic covariance matrices is obtained.

In Section 4 we apply the results to the problem of hypothesis testing for null effects. We can test a scale-invariant null hypothesis. In particular, we can determine which coordinates of the predictor have significant effects on the response.

In Section 5 we discuss some possible generalization of our methods to other settings. One case involves the presence of a stratification variable, such as sex, or district, or age, as part of the predictor variable. We also discuss the general dimension reduction model (1.2).

In Section 6 we address the issue of visualizing the data. We treat this as a dimension reduction problem and argue that in some cases it may be sufficient to plot $\mathbf{Y}$ against $\beta'\mathbf{u}$. A simulation study is reported to demonstrate the effectiveness of our approach.

In Section 7 we briefly indicate one way to estimate the constant of proportionality if the link function g is known. Section 8 contains the results of our techniques applied to a binary regression example. The Appendix presents some technical details.

2. BASIC THEORY

2.1 Preliminaries

In model (1.1), suppose that instead of $\mathbf{x}$ we can observe only a surrogate $\mathbf{w}$, which is related to $\mathbf{x}$ via the linear model

$$\mathbf{w} = \gamma + \Gamma\mathbf{x} + \delta, \tag{2.1}$$

where Γ is a q by p matrix, which can be known, unknown, or partly known. We assume that δ is independent of $\mathbf{x}$ and ε, although this can be relaxed to δ and $\mathbf{Y}$ independent given $\beta'\mathbf{x}$. An important case is when $p = q$ and Γ equals the identity.

We shall present some methods of estimating β, for $p \geq 2$, up to a constant of proportionality *without knowledge of the functional form of g*. Similar to sliced inverse regression (Li 1991), the key idea in our approach is to consider the inverse regression curve $\eta(y) = E(\mathbf{w}|\mathbf{Y} = y)$. Theorem 2.1 shows that this curve will fall on a straight line under the following condition: For any direction b in R^p, $E(b'\mathbf{x}|\beta'\mathbf{x})$ is linear in $\beta'\mathbf{x}$; that is,

$$E(b'\mathbf{x}|\beta'\mathbf{x}) = c_0 + c_1\beta'\mathbf{x} \tag{2.2}$$

for some constants c_0, c_1. An important special case for (2.2) is when the distribution of $\mathbf{x}$ is elliptically symmetric. But as discussed in Li (1991), (2.2) is much weaker than elliptic symmetry because it need be satisfied only by the vector β. See Remark 2.2 for more discussion.

Theorem 2.1. Under (1.1), (2.1), and (2.2), we have

$$\eta(y) = E(\mathbf{w}|\mathbf{Y} = y) = E(\mathbf{w}) + c(y)\Gamma\Sigma_{\mathbf{x}}\beta,$$

where the scalar function $c(y)$ equals $(\beta'\Sigma_{\mathbf{x}}\beta)^{-1}E(\beta'(\mathbf{x} - E\mathbf{x})|\mathbf{Y} = y)$.

Proof. We can assume that $E\mathbf{w} = E\mathbf{x} = 0$. It may be shown that (2.2) implies that $E(\mathbf{x}|\beta'\mathbf{x}) = c_1(\beta'\mathbf{x})\Sigma_{\mathbf{x}}\beta$, where $c_1(\beta'\mathbf{x}) = (\beta'\Sigma_{\mathbf{x}}\beta)^{-1}\beta'\mathbf{x}$. Then, by conditioning, we find that

$$\begin{aligned} E(\mathbf{w}|\mathbf{Y}) &= E\{E(\mathbf{w}|\beta'\mathbf{x}, \mathbf{Y})|\mathbf{Y}\} \\ &= E\{\Gamma E(\mathbf{x}|\beta'\mathbf{x})|\mathbf{Y}\} = c(\mathbf{Y})\Gamma\Sigma_{\mathbf{x}}\beta, \end{aligned}$$

where $c(y) = E\{c_1(\beta'\mathbf{x})|\mathbf{Y} = y\}$. This proves the theorem.

There are two ways to apply this theorem, as described in Sections 2.2 and 2.3.

2.2 Linear Regression-Type Method

Suppose that a standard linear least squares regression of $\mathbf{Y}$ against $\mathbf{w}$ is conducted:

$$\min_{b_1 \in R^q, a \in R} E(\mathbf{Y} - a - b_1'\mathbf{w})^2. \tag{2.3}$$

Then the regression slope b_1 satisfies

$$\begin{aligned} b_1 &= \Sigma_{\mathbf{w}}^{-1}E\mathbf{Y}(\mathbf{w} - E\mathbf{w}) = c\Sigma_{\mathbf{w}}^{-1}\Gamma\Sigma_{\mathbf{x}}\beta \\ &= (\beta'\Sigma_{\mathbf{x}}\beta)^{-1}\text{cov}(\beta'\mathbf{x}, \mathbf{Y})\Sigma_{\mathbf{w}}^{-1}\Gamma\Sigma_{\mathbf{x}}\beta. \end{aligned} \tag{2.4}$$

If $g(\beta'\mathbf{x}, \varepsilon) = \alpha + \beta'\mathbf{x} + \varepsilon$, then the constant c equals 1.

The proof of (2.4) follows directly from Theorem 2.1. Now recall that

$$\mathbf{u} = L\mathbf{w} = \text{cov}(\mathbf{x}, \mathbf{w})\Sigma_{\mathbf{w}}^{-1}\mathbf{w} = \Sigma_{\mathbf{x}}\Gamma'\Sigma_{\mathbf{w}}^{-1}\mathbf{w}. \tag{2.5}$$

Plugging (2.4) and (2.5) into the minimization problem (2.3), we see that β is proportional to the solution, b_{ls}, of the following:

$$\min_{b \in R^p, a \in R} E(\mathbf{Y} - a - b'\mathbf{u})^2. \tag{2.6}$$

This result can be viewed as an extension of Brillinger (1977, 1983), when the usual linear least squares estimate was shown to be consistent for estimating the direction of the slope vector when the regressor variable is free of error. Li and Duan

(1989) extended Brillinger's result to other regression estimates, including generalized linear models and M estimates.

2.3 SIR-Type Method

Because the regression curve $\eta(y) = E(\mathbf{w}|\mathbf{Y} = y)$ falls on a straight line, we can use the same method as SIR (Li 1991) to suggest an estimate. It is more convenient for us to consider the curve $\zeta(y) = E(\mathbf{u}|\mathbf{Y} = y)$, where $\mathbf{u}$ is defined by (2.5). From Theorem 2.1 we find that

$$\zeta(y) = E(\mathbf{u}) + c(y)\boldsymbol{\Sigma}_{\mathbf{u}}\beta, \tag{2.7}$$

where, by (2.5), $\boldsymbol{\Sigma}_{\mathbf{u}} = \boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{x}}$ denotes the covariance matrix of $\mathbf{u}$. Thus, in terms of $\mathbf{u}$, the inverse regression curve is a straight line with the direction specified in a way exactly the same as in the regressor-error-free case discussed by Li (1991, Theorem 3.1). Denote the covariance matrix $\operatorname{cov}\{\zeta(\mathbf{Y})\}$ by $\boldsymbol{\Sigma}_{\zeta}$. We can find the nondegenerate direction by applying a suitable principal component analysis, as the following corollary suggests.

Corollary 2.1. Under the same conditions as given in Theorem 2.1, the covariance matrix $\boldsymbol{\Sigma}_{\zeta}$ has rank at most 1. Assuming that $\boldsymbol{\Sigma}_{\zeta}$ is of rank 1, let v be the nonzero eigenvector for the eigenvalue decomposition of $\boldsymbol{\Sigma}_{\zeta}$ with respect to $\boldsymbol{\Sigma}_{\mathbf{u}}$: $\boldsymbol{\Sigma}_{\zeta}v = \lambda\boldsymbol{\Sigma}_{\mathbf{u}}v$, where λ is the nonzero eigenvalue. Then v is proportional to β: for some scalar c, $v = c\beta$.

Note that the eigenvalue $\lambda = E[E\{\beta'(\mathbf{x} - E\mathbf{x})|\mathbf{Y}\}^2]/\beta'\boldsymbol{\Sigma}_{\mathbf{u}}\beta$.

Remark 2.1. As mentioned previously, in (2.1) we require only that conditional on $\beta'\mathbf{x}$, the distribution of $\mathbf{Y}$ is independent of δ.

Remark 2.2. Diaconis and Freedman (1984) showed that almost all low-dimensional projections of high-dimensional data are approximately normal. Hall and Li (1993) showed that from a Bayesian perspective, if β assumes a vague prior distribution, then as the dimensionality tends to ∞, (2.2) holds approximately with probability approaching 1. This implies that our assumption (2.2) is realistic for many high-dimensional data sets. On the other hand, if (2.2) is severely violated for some direction b, then it may be difficult to determine which direction in the space spanned by b and β is truly responsible for determining $\mathbf{Y}$, given a realistic sample size. Li (1990a) demonstrated how SIR can help find the space spanned by the true direction β and the direction of b for which (2.2) is most severely violated. We expect a similar extension to our case; namely, the first two (or more, if necessary) eigenvectors of SIR may help recover the most severe nonlinearity in the design distribution.

Remark 2.3. In the regressor-error-free case, there are some interesting methods that do not require the design condition (2.2) to estimate β consistently up to a constant of proportionality; see for example, the average derivative method of Härdle and Stoker (1989) and some variants of projection pursuit regression as given in Hall (1989) and Chen (1991). These methods use the property that the response function $E(\mathbf{Y}|\mathbf{x})$ depends on $\mathbf{x}$ only through $\beta'\mathbf{x}$. Extension of these techniques to measurement error models would be of interest.

Remark 2.4. A necessary and sufficient condition for the matrix $\boldsymbol{\Sigma}_{\zeta}$ to be nonzero is that $\operatorname{cov}\{h(\mathbf{Y}), \beta'\mathbf{x}\} \neq 0$, for some transformation $h(\cdot)$. This holds for most models where $E(\mathbf{Y}|\beta'\mathbf{x})$ is strictly monotone or for heteroscedastic models such as $\mathbf{Y} = \varepsilon \cdot g(\beta'\mathbf{x})$.

Remark 2.5. It is easy to see that each row of the linear transformation L is the regression slope vector for linearly regressing the corresponding coordinate of $\mathbf{x}$ against $\mathbf{w}$, intercept included. In particular, the regression slope for the variable $\beta'\mathbf{x}$ against $\mathbf{w}$ equals $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\operatorname{cov}(\mathbf{w}, \mathbf{x}'\beta) = L'\beta$. Therefore, the variable $\beta'L\mathbf{w} = \beta'\mathbf{u}$ has higher correlation with $\beta'\mathbf{x}$ than any other linear combination of $\mathbf{w}$: $\operatorname{corr}(\beta'\mathbf{x}, \beta'\mathbf{u}) \geq \operatorname{corr}(\beta'\mathbf{x}, b'\mathbf{w})$.

Remark 2.6. If the joint distribution of $\mathbf{x}$ and $\mathbf{w}$ is normal, we can extend the result of Li and Duan (1989) to the error-in-regressors problem by pretending that we have observed an error-free regressor $\mathbf{u}$. Specifically, for any function $\rho(\mathbf{Y}, \theta)$ that is convex in θ, under (1.1) the solution (a_ρ, b_ρ) of the minimization

$$\min_{a\in R, b\in R^p} E\rho(\mathbf{Y}, a + b'\mathbf{u})$$

satisfies the condition that b_ρ is proportional to β. The proof of this result is given in Appendix A. In fact, the normality assumption can be weakened by assuming (2.1) and that, conditional on $\beta'\mathbf{u}$, $\beta'\mathbf{x}$ is independent of $\mathbf{u}$.

3. ESTIMATION

Given an iid sample, $(\mathbf{Y}_i, \mathbf{w}_i)$, $i = 1, \ldots, n$, we discuss how to implement the two methods described in Section 2. As in (2.5), define

$$L = \boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} = \operatorname{cov}(\mathbf{x}, \mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}.$$

We distinguish among four different situations:

1. L is known.
2. L is unknown and estimated by an independent validation sample of $(\mathbf{x}, \mathbf{w})$, which is assumed to be representative in the sense that the covariance of $\mathbf{x}$ is the same in the primary and validation data sets.
3. L is unknown, but $p = q$ and $\boldsymbol{\Gamma}$ is known in (2.1), so that L can be estimated by an independent representative sample containing replicates of $\mathbf{w}$.
4. Either independent validation or replication samples are taken, but they are not representative because the covariance of $\mathbf{x}$ differs from that in the primary data set.

The case where L is known is important primarily to set up the theory, although there may be situations where this case occurs. In addition, sensitivity analyses for different L's may be of interest when there is no additional information about the relationship between $\mathbf{x}$ and $\mathbf{w}$. This case is treated in Sections 3.1 and 3.2.

The use of validation is treated in Sections 3.3 and 3.4. We assume there and in Sections 3.5 and 3.6 that the in-

dependent data set is *representative;* one such example is discussed by Rosner et al. (1989).

In many instances, Γ may be known. This occurs especially when $\mathbf{w}$ is an unbiased estimate of $\mathbf{x}$, the most common case occurring in the literature (Fuller 1987). For this important situation, to estimate L we show in Sections 3.5 and 3.6 that it is only necessary to have replicates $\mathbf{w}_{i1}$ and $\mathbf{w}_{i2}$ of $\mathbf{x}_i$ for some i.

The general case of possibly nonrepresentative validation or replication is treated in Sections 3.7 and 3.8. We will consider only the case that the replicated data set is independent of the primary data set, although other cases are possible.

3.1 Least Squares With Known L

If L is known, define $\mathbf{u}_i = L\mathbf{w}_i$. The least squares estimate for (2.6) is

$$\tilde{b}_{\text{ls}} = \tilde{\mathbf{\Sigma}}_{\mathbf{u}}^{-1}(n-1)^{-1}\sum_{i=1}^{n}\mathbf{Y}_i(\mathbf{u}_i - \bar{\mathbf{u}}), \tag{3.1}$$

where $\tilde{\mathbf{\Sigma}}_{\mathbf{u}} = L\hat{\mathbf{\Sigma}}_{1\mathbf{w}}L'$ is the sample covariance for $\mathbf{u}$ and $\hat{\mathbf{\Sigma}}_{1\mathbf{w}}$ is the sample covariance of $\mathbf{w}$. Root n consistency and asymptotic normality of this estimate follows easily. Define the ith theoretical residual after regression: $e_i = \mathbf{Y}_i - E\mathbf{Y} - b_{\text{ls}}'(\mathbf{u}_i - E\mathbf{u})$. Then we have the expansion

$$\tilde{b}_{\text{ls}} = b_{\text{ls}} + n^{-1}\sum_{i=1}^{n}\mathbf{\Sigma}_{\mathbf{u}}^{-1}e_i(\mathbf{u}_i - E\mathbf{u}) + O_p(n^{-1}). \tag{3.2}$$

Clearly, $\tilde{b}_{\text{ls}}$ is asymptotically normal, with mean b_{ls} defined by (2.6) and covariance matrix

$$\mathbf{\Sigma}(\tilde{b}_{\text{ls}}) = n^{-1}\mathbf{\Sigma}_{\mathbf{u}}^{-1}\text{cov}\{e_i(\mathbf{u}_i - E\mathbf{u})\}\mathbf{\Sigma}_{\mathbf{u}}^{-1}. \tag{3.3}$$

3.2 SIR With Known L

The SIR-type estimate can be constructed as in Li (1991):

1. Divide the range of $\mathbf{Y}$ into H slices and let $\hat{p}_h$ be the proportion of $\mathbf{Y}_i$'s falling into the hth slice I_h.
2. Within each slice compute the sample mean of $\mathbf{u}$, $\bar{\mathbf{u}}_h = (n\hat{p}_h)^{-1}\mathbf{\Sigma}_{\mathbf{Y}_i\in I_h}\mathbf{u}_i$, $h = 1, \ldots, H$.
3. Form the covariance matrix $\tilde{\mathbf{\Sigma}}_{\xi} = \mathbf{\Sigma}_{h=1}^{H}\hat{p}_h(\bar{\mathbf{u}}_h - \bar{\mathbf{u}})(\bar{\mathbf{u}}_h - \bar{\mathbf{u}})'$. Then conduct an eigenvalue decomposition of $\tilde{\mathbf{\Sigma}}_{\xi}$ with respect to the sample covariance of $\mathbf{u}$, $\tilde{\mathbf{\Sigma}}_{\mathbf{u}}$:

$$\tilde{\mathbf{\Sigma}}_{\xi}\tilde{b}_{\text{SIR}} = \tilde{\lambda}\tilde{\mathbf{\Sigma}}_{\mathbf{u}}\tilde{b}_{\text{SIR}},$$

where $\tilde{\lambda}$ is the largest eigenvalue.

The asymptotic distribution of $\tilde{b}_{\text{SIR}}$ follows from arguments similar to Li (1991) or Duan and Li (1991) for fixed H, and arguments similar to Hsing and Carroll (1992) for $H = n/2$. The fixed H case is given in detail in Theorem 3.1 below.

Due to the affine invariance, we can without loss of generality assume that $E\mathbf{w} = 0$. Recall the definition of $c(y)$ from Theorem 2.1 and define $\mathbf{k} = [E\{c(\mathbf{Y})|\mathbf{Y}\in I_1\}, \ldots, E\{c(\mathbf{Y})|\mathbf{Y}\in I_H\}]' = (k_1, \ldots, k_H)'$, $\bar{\mathbf{U}} = (\bar{\mathbf{u}}_1, \ldots, \bar{\mathbf{u}}_H)$, and $\hat{D}$ = a diagonal matrix with diagonal elements $\hat{p}_1, \ldots, \hat{p}_H$. Then by the law of large numbers, the q by H matrix $\bar{\mathbf{U}}$ converges to $E\bar{\mathbf{U}} = \mathbf{\Sigma}_{\mathbf{u}}\beta\mathbf{k}'$ (more precisely, the expectation is conditional on $\hat{p}_h \neq 0$) at root n rate. Let $\Delta_{\mathbf{u}} = \bar{\mathbf{U}} - E\bar{\mathbf{U}}$. Then

$$\begin{aligned}\tilde{\mathbf{\Sigma}}_{\xi} &= (E\bar{\mathbf{U}} + \Delta_{\mathbf{u}})\hat{D}(E\bar{\mathbf{U}} + \Delta_{\mathbf{u}})' \\ &= \mathbf{k}'\hat{D}\mathbf{k}\mathbf{\Sigma}_{\mathbf{u}}\beta\beta'\mathbf{\Sigma}_{\mathbf{u}} + \Delta_{\mathbf{u}}\hat{D}\mathbf{k}\beta'\mathbf{\Sigma}_{\mathbf{u}} + \mathbf{\Sigma}_{\mathbf{u}}\beta\mathbf{k}'\hat{D}\Delta_{\mathbf{u}}' + O_p(n^{-1}) \\ &= \mathbf{k}'\hat{D}\mathbf{k}\{\mathbf{\Sigma}_{\mathbf{u}}\beta + (\mathbf{k}'\hat{D}\mathbf{k})^{-1}\Delta_{\mathbf{u}}\hat{D}\mathbf{k}\} \\ &\quad \times \{\mathbf{\Sigma}_{\mathbf{u}}\beta + (\mathbf{k}'\hat{D}\mathbf{k})^{-1}\Delta_{\mathbf{u}}\hat{D}\mathbf{k}\}' + O_p(n^{-1}).\end{aligned} \tag{3.4}$$

It follows immediately that the largest eigenvector $\hat{b}_{\text{sir}}$ is approximately proportional to

$$\tilde{\mathbf{\Sigma}}_{\mathbf{u}}^{-1}(\mathbf{k}'\hat{D}\mathbf{k}\mathbf{\Sigma}_{\mathbf{u}}\beta + \Delta_{\mathbf{u}}\hat{D}\mathbf{k}). \tag{3.5}$$

We can relate (3.5) to the least squares approximation (3.2) by considering the transformed variables $\tilde{\mathbf{Y}}_i = k_h$ if $\mathbf{Y}_i$ falls in the hth slice; that is,

$$\tilde{\mathbf{Y}}_i = \sum_{h=1}^{H}\xi_h(i)k_h; \qquad \xi_h(i) = I(Y_i \in h\text{th slice}).$$

A straightforward derivation as outlined in Appendix B shows that (3.5) can be approximated by

$$\tilde{\mathbf{\Sigma}}_{\mathbf{u}}^{-1}n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{Y}}_i(\mathbf{u}_i - \bar{\mathbf{u}}), \tag{3.6}$$

which would be equal to (3.1) approximately if we had transformed $\mathbf{Y}_i$ to $\tilde{\mathbf{Y}}_i$. Therefore, we can apply (3.2) to our case with e_i replaced by $\tilde{e}_i = \tilde{\mathbf{Y}}_i - E\tilde{\mathbf{Y}}_i - b_{\text{sir}}'(\mathbf{u}_i - E\mathbf{u})$, where b_{sir} is the regression slope of $\tilde{\mathbf{Y}}$ against $\mathbf{u}$. Because (3.5) converges to $\mathbf{k}'D\mathbf{k}\beta$, we see that

$$b_{\text{sir}} = \mathbf{k}'D\mathbf{k}\beta. \tag{3.7}$$

There is a slight ambiguity for the definition of the eigenvector $\tilde{b}_{\text{sir}}$ in Step 3, because any nonzero multiple of an eigenvector is also an eivenvector. Because we are using $\tilde{b}_{\text{sir}}$ to estimate the direction of β, it does not matter which version is used.

Theorem 3.1. Under the same conditions as given in Theorem 2.1, the estimate $\tilde{b}_{\text{sir}}$ is root n consistent in estimating the direction of β. Moreover, there exists a version of $\tilde{b}_{\text{sir}}$ that is asymptotically normal with mean b_{sir} given by (3.7) and covariance matrix

$$\mathbf{\Sigma}(\tilde{b}_{\text{sir}}) = n^{-1}\mathbf{\Sigma}_{\mathbf{u}}^{-1}\text{cov}\{\tilde{e}_i(\mathbf{u}_i - E\mathbf{u})\}\mathbf{\Sigma}_{\mathbf{u}}^{-1}.$$

Remark 3.1. We can choose the version $\tilde{b}_{\text{sir}}$ with unit length (with respect to $\mathbf{\Sigma}_{\mathbf{u}}$ or the identity matrix, for example) and the largest coordinate value of which is positive. The asymptotic distribution for this version can be easily derived from Theorem 3.1 and the following lemma, which is proved in Appendix C.

Lemma 3.1. For any sequence of random vectors, $\mathbf{z}_i$, such that $n^{1/2}(\mathbf{z}_i - \mu)$ converges to a normal with mean vector 0 and covariance matrix $\mathbf{\Sigma}$, the sequence of normalized random vectors $n^{1/2}(\mathbf{z}_i/\|\mathbf{z}_i\|^{-1} - \mu/\|\mu\|)$ converges to a normal distribution with mean 0 and asymptotic covariance matrix $\|\mu\|^{-2}P_2\mathbf{\Sigma}P_2$, where P_2 is the projection matrix $I - (\|\mu\|)^{-2}\mu\mu'$.

3.3 Least Squares With Unknown L: Representative Validation

Most often, L is unknown and has to be estimated. In this section we consider the use of an independent *external validation* sample of size m; that is, an independent data set consisting of the measurements $(\mathbf{x}_i, \mathbf{w}_i)$, $i = n+1, \ldots, n+m$. Typically, m is smaller than n. We will assume for this purpose that $\mathbf{x}$ has the same covariance in the primary as well as the validation data; the case that it does not is covered in Section 3.7.

We now study the effect on our estimates in (3.1) due to the uncertainty from estimating L. We consider the estimate of L defined by

$$\hat{L} = \widehat{\mathrm{cov}}(\mathbf{x}, \mathbf{w})\hat{\boldsymbol{\Sigma}}_{2\mathbf{w}}^{-1}, \tag{3.8}$$

where $\widehat{\mathrm{cov}}(\mathbf{x}, \mathbf{w})$ is the sample covariance between $\mathbf{x}$ and $\mathbf{w}$ and $\hat{\boldsymbol{\Sigma}}_{2\mathbf{w}}$ is the sample covariance of $\mathbf{w}$, all based on the validation sample $(\mathbf{x}_i, \mathbf{w}_i)$, $i = n+1, \ldots, n+m$. Each row of $\hat{L}$ is the usual least squares regression slope of the corresponding coordinate of $\mathbf{x}$ against $\mathbf{w}$.

Denote $\hat{\mathbf{u}}_i = \hat{L}\mathbf{w}_i$ for $i = 1, \ldots, n$ and define the associated sample covariance

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \hat{L}\hat{\boldsymbol{\Sigma}}_{1\mathbf{w}}\hat{L}'. \tag{3.9}$$

We shall replace $\mathbf{u}_i$'s by $\hat{\mathbf{u}}_i$'s in constructing our estimates. The resulting estimates of $\boldsymbol{\beta}$ will be denoted by $\hat{b}_{\mathrm{ls}}$ and $\hat{b}_{\mathrm{sir}}$.

The least squares estimate $\hat{b}_{\mathrm{ls}}$ is obtained from (3.1) by replacing $\mathbf{u}_i$ by $\hat{\mathbf{u}}_i$:

$$\hat{b}_{\mathrm{ls}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1}(n-1)^{-1}\sum_{i=1}^{n}\mathbf{Y}_i(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}}),$$

where $\bar{\hat{\mathbf{u}}}$ denotes the sample mean of $\hat{\mathbf{u}}_i$'s. A standard expansion shows that, no matter how L is estimated,

$$\hat{b}_{\mathrm{ls}} - b_{\mathrm{ls}} = \boldsymbol{\Sigma}_{\mathbf{u}}^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{u}_i - E\mathbf{u})e_i - \boldsymbol{\Sigma}_{\mathbf{u}}^{-1}L\boldsymbol{\Sigma}_{\mathbf{w}}(\hat{L} - L)'b_{\mathrm{ls}} + O_p(n^{-1} + m^{-1}). \tag{3.10}$$

Now let

$$\mathbf{r}_i = \mathbf{x}_i - \mathbf{u}_i = \mathbf{x}_i - L\mathbf{w}_i; \qquad \boldsymbol{\Lambda}_{\mathbf{rw}i} = (\mathbf{r}_i - E\mathbf{r}_i)(\mathbf{w}_i - E\mathbf{w})'.$$

A standard expansion for least squares gives

$$\hat{L} - L = m^{-1}\sum_{i=n+1}^{n+m}\boldsymbol{\Lambda}_{\mathbf{rw}i}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + O_p(m^{-1}). \tag{3.11}$$

Hence we obtain the expansion

$$\hat{b}_{\mathrm{ls}} = b_{\mathrm{ls}} - m^{-1}\sum_{i=n+1}^{n+m}\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}L\boldsymbol{\Lambda}_{\mathbf{rw}i}'b_{\mathrm{ls}} + \boldsymbol{\Sigma}_{\mathbf{u}}^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{u}_i - E\mathbf{u})e_i + O_p(m^{-1}). \tag{3.12}$$

Note that the last two terms are independent. It is now clear that $\hat{b}_{\mathrm{ls}}$ is consistent. Compared with (3.2), we see that the cost of estimating L is the presence of the second term in (3.12) from the validation sample.

Theorem 3.2. The asymptotic covariance matrix of $\hat{b}_{\mathrm{ls}}$ is given by

$$\boldsymbol{\Sigma}(\hat{b}_{\mathrm{ls}}) = \mathrm{cov}\{I(b_{\mathrm{ls}})\} + \boldsymbol{\Sigma}(\tilde{b}_{\mathrm{ls}}),$$

where $\boldsymbol{\Sigma}(\tilde{b}_{\mathrm{ls}})$ is given by (3.3) and

$$\mathrm{cov}\{I(b_{\mathrm{ls}})\} = m^{-1}\mathrm{cov}\{\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}(\mathbf{u}_i - E\mathbf{u})(\mathbf{r}_i - E\mathbf{r}_i)b_{\mathrm{ls}}'\}.$$

3.4 SIR With Unknown L: Representative Validation

After estimating L by $\hat{L}$, the SIR-type estimate can be carried out in the same way as described in Section 3.2. The matrix $\tilde{\boldsymbol{\Sigma}}_\zeta$ will be replaced by

$$\hat{\boldsymbol{\Sigma}}_\zeta = \hat{L}\hat{\boldsymbol{\Sigma}}_\eta\hat{L}', \tag{3.13}$$

where

$$\hat{\boldsymbol{\Sigma}}_\eta = \sum_{h=1}^{H}\hat{p}_h(\bar{\mathbf{w}}_h - \bar{\mathbf{w}})(\bar{\mathbf{w}}_h - \bar{\mathbf{w}})', \tag{3.14}$$

and where $\bar{\mathbf{w}}_h$ is the slice mean for $\mathbf{w}$, $(n\hat{p}_h)^{-1}\sum_{Y_i \in I_h}\mathbf{w}_i$. The maximum eigenvector of the eigenvalue decomposition $\hat{\boldsymbol{\Sigma}}_\zeta$ with respect to $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ [see (3.9)] is our estimate $\hat{b}_{\mathrm{sir}}$. We shall find the asymptotic distribution for $\hat{b}_{\mathrm{sir}}$ as follows.

First, following the same derivation as the one that leads to (3.4), we obtain

$$\hat{\boldsymbol{\Sigma}}_\eta \doteq \mathbf{k}'\hat{D}\mathbf{k}\{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\beta} + (\mathbf{k}'\hat{D}\mathbf{k})^{-1}\Delta_{\mathbf{w}}\hat{D}\mathbf{k}\} \times \{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\beta} + (\mathbf{k}'\hat{D}\mathbf{k})^{-1}\Delta_{\mathbf{w}}\hat{D}\mathbf{k}\}',$$

where $\Delta_{\mathbf{w}} = (\bar{\mathbf{w}}_1 - E\bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_H - E\bar{\mathbf{w}}_H)$. Based on this, from (3.13) we approximate $\hat{\boldsymbol{\Sigma}}_\zeta$ and find that one version of the eigenvector $\hat{b}_{\mathrm{sir}}$ approximately equals

$$\hat{b}_{\mathrm{sir}} \doteq \hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1}\hat{L}(\mathbf{k}'\hat{D}\mathbf{k}\cdot\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\beta} + \Delta_{\mathbf{w}}\hat{D}\mathbf{k}). \tag{3.15}$$

Following an argument similar to (A.2), we can approximate (3.15) by

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1}(n-1)^{-1}\sum_{i=1}^{n}\tilde{\mathbf{Y}}_i(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}}). \tag{3.16}$$

Comparing (3.16) with (3.12), we see that, asymptotically, $\hat{b}_{\mathrm{sir}}$ would be equivalent to the least squares estimate of Section 3.3 if we had transformed $\mathbf{Y}_i$ to $\tilde{\mathbf{Y}}_i$. Hence we can apply Theorem 3.2 to obtain the asymptotic distribution of $\hat{b}_{\mathrm{sir}}$.

Theorem 3.3. There exists a version of $\hat{b}_{\mathrm{sir}}$ with the asymptotic mean b_{sir} given by (3.7) and the covariance matrix

$$\boldsymbol{\Sigma}(\hat{b}_{\mathrm{sir}}) = \mathrm{cov}\{I(b_{\mathrm{sir}})\} + \boldsymbol{\Sigma}(\tilde{b}_{\mathrm{sir}}),$$

where $\boldsymbol{\Sigma}(\tilde{b}_{\mathrm{sir}})$ is given in Theorem 3.1, and $\mathrm{cov}\{I(b_{\mathrm{sir}})\}$ is the same as the term $\mathrm{cov}\{I(b_{\mathrm{ls}})\}$ given in Theorem 3.2, with b_{ls} replaced by b_{sir}.

Remark 3.2. Note that the term $\mathrm{cov}\{I(b_{\mathrm{sir}})\}$ (respectively, $\mathrm{cov}\{I(b_{\mathrm{ls}})\}$) is the asymptotic covariance matrix for the estimated slope when we regress $b_{\mathrm{sir}}'\mathbf{x}$ (respectively, $b_{\mathrm{ls}}'\mathbf{x}$) against $\mathbf{u}$ based on the validation sample, if b_{sir} (respectively, b_{ls}) were known. Thus the additional uncertainty in our estimate due to estimating L is easy to assess. This information may be particularly useful in planning of the sample sizes m and n.

Remark 3.3. In Theorem 3.3 we assumed that the slices are fixed. In practice it is more convenient to choose approximately the same number of observations for each slice, unless $\mathbf{Y}$ is discrete. This makes our procedure invariant under monotone transformations of $\mathbf{Y}$. The case where H increases as the sample size increases (in particular two observations per slice) can be treated using the methods of Hsing and Carroll (1992).

3.5 Least Squares With Unknown L: Representative Replication

An important special case occurs when Γ is known and $p = q$. Without loss we will take $\Gamma = I$, in which case $\mathbf{w}$ is an unbiased surrogate for $\mathbf{x}$. In many experiments, instead of validation we will have a replicated data set; that is,

$$\mathbf{w}_{ij} = \alpha + \mathbf{x}_i + \delta_{ij}, \qquad j = 1, 2; \qquad i = n+1, \ldots, n+m. \quad (3.17)$$

If Σ_δ is the covariance of δ_{ij}, then we find that $L = I - \Sigma_\delta \Sigma_{\mathbf{w}}^{-1}$. Define $\hat{\Sigma}_\delta$ and $\hat{\Sigma}_{\mathbf{w}} - (1/2)\hat{\Sigma}_\delta$ to be the sample covariance matrices of $(\mathbf{w}_{i1} - \mathbf{w}_{i2})/2^{1/2}$ and $(\mathbf{w}_{i1} + \mathbf{w}_{i2})/2$ and define

$$\hat{L} = I - \hat{\Sigma}_\delta \hat{\Sigma}_{\mathbf{w}}^{-1}. \quad (3.18)$$

Then, while we prefer the method of Section 3.7, all the results of Section 3.3 hold if we replace Λ_{rwi} by

$$\Lambda_i = \left(\frac{1}{4}\right)\left[\Sigma_\delta \Sigma_{\mathbf{w}}^{-1}\left\{(\mathbf{w}_{i1} + \mathbf{w}_{i2} - 2E\mathbf{w})(\mathbf{w}_{i1} + \mathbf{w}_{i2} - 2E\mathbf{w})'\right\} + (\Sigma_\delta \Sigma_{\mathbf{w}}^{-1} - 2I)(\mathbf{w}_{i1} - \mathbf{w}_{i2})(\mathbf{w}_{i1} - \mathbf{w}_{i2})'\right]. \quad (3.19)$$

3.6 SIR With Unknown L: Representative Replication

In the replication model (3.17), with the estimate (3.18), the results of Section 3.4 go through with the only change that $\Lambda_{\mathbf{rw}i}$ is replaced by (3.19).

3.7 Least Squares With Unknown L: General Case

Formula (2.5) refers to the primary sample. In many problems the classical additive measurement error model (2.1) can be assumed to hold both for the primary and the validation/replication data sets, with the same values of (γ, Γ) and the covariance matrix of δ; see Carroll (1989) for discussion. But there are important instances where the marginal distribution of $\mathbf{x}$ is not the same in the primary and validation/replication data sets, in which case adjustments must be made. Thus we will estimate (Γ, Σ_δ) from the validation/replication data sets and then use the primary data set to estimate $\Sigma_{\mathbf{x}}$.

Assume that $\Omega = \Gamma'\Gamma$ is of full rank. Then $\Sigma_{\mathbf{x}} = \Omega^{-1}\Gamma'(\Sigma_{\mathbf{w}} - \Sigma_\delta)\Gamma\Omega^{-1}$. If Γ is unknown, let $\hat{\Gamma}$ be the least squares estimate; otherwise, set $\hat{\Gamma} = \Gamma$. Let $\hat{\Sigma}_\delta$ be an estimate of Σ_δ, which we will assume to have the expansion

$$\hat{\Sigma}_\delta - \Sigma_\delta = m^{-1} \sum_{i=n+1}^{n+m} \Delta_i + O_p(m^{-1}).$$

In the case of validation, $\Delta_i = \delta_i\delta_i' - \Sigma_\delta$. For replication, $\Delta_i = \delta_{i*}\delta_{i*}' - \Sigma_\delta$, where $\delta_{i*} = (\mathbf{w}_{i1} - \mathbf{w}_{i2})/2^{1/2}$.

Now define $\hat{L} = \hat{\Sigma}_{\mathbf{x}}\hat{\Gamma}'\hat{\Sigma}_{1\mathbf{w}}^{-1}$, where $\hat{\Sigma}_{\mathbf{x}} = \hat{\Omega}^{-1}\hat{\Gamma}'(\hat{\Sigma}_{1\mathbf{w}} - \hat{\Sigma}_\delta)\hat{\Gamma}\hat{\Omega}^{-1}$. For $i = 1, \ldots, n$ define $s_i = (\mathbf{w}_i - E\mathbf{w})(\mathbf{w}_i - E\mathbf{w})' - \Sigma_{\mathbf{w}}$, $l_i = \Omega^{-1}\Gamma' s_i \Gamma\Omega^{-1}\Gamma'\Sigma_{\mathbf{w}}^{-1} - \hat{L}s_i\Sigma_{\mathbf{w}}^{-1}$ and

$$\mathcal{A}_i = \Sigma_{\mathbf{u}}^{-1}\{(\mathbf{u}_i - E\mathbf{u})e_i - L\Sigma_{\mathbf{w}}l_i'b_{\mathrm{ls}}\}.$$

For $i = n+1, \ldots, n+m$, under replication with $\Gamma = I$ define $\mathcal{B}_i = 0$; for validation define $\mathcal{B}_i = \Sigma_{\mathbf{x}}^{-1}(\mathbf{x}_i - E\mathbf{x})\delta_i'$. Further define

$$\tau_i = (-\Omega^{-1}\Gamma'\Delta_i\Gamma\Omega^{-1}\Gamma' + \Sigma_{\mathbf{x}}\mathcal{B}_i - \Omega^{-1}\Gamma'\mathcal{B}_i\Sigma_{\mathbf{x}}\Gamma' - \Sigma_{\mathbf{x}}\mathcal{B}_i'\Gamma\Omega^{-1}\Gamma')\Sigma_{\mathbf{w}}^{-1}$$

and $\mathcal{A}_i = \Sigma_{\mathbf{u}}^{-1}L\Sigma_{\mathbf{w}}\tau_i'b_{\mathrm{ls}}$. Then calculations outlined in Appendix D show that

$$\hat{b}_{\mathrm{ls}} - b_{\mathrm{ls}} = n^{-1}\sum_{i=1}^{n}\mathcal{A}_i + m^{-1}\sum_{i=n+1}^{n+m}\mathcal{A}_i + O_p(n^{-1} + m^{-1}). \quad (3.20)$$

Equation (3.20) allows us to state the following result:

Theorem 3.4. $\hat{b}_{\mathrm{ls}}$ is asymptotically normally distributed with mean b_{ls} and covariance matrix

$$n^{-1}\mathrm{cov}(\mathcal{A}_1) + m^{-1}\mathrm{cov}(\mathcal{A}_{n+m}). \quad (3.21)$$

3.8 SIR With Unknown L: General Case

This may be treated in the same way as in Section 3.7; replace b_{ls} by b_{sir} and replace $\mathbf{Y}_i$ by $\tilde{\mathbf{Y}}_i$.

4. STATISTICAL INFERENCE

We shall show how to apply the results of Section 3 to hypothesis testing for null effects. Hypothesis testing has not been much discussed in the nonlinear measurement error model literature. An exception is the case of testing for a simultaneous null effect in all components of $\mathbf{x}$ measured with error, where score test ideas can be used (see Stefanski and Carroll 1990; Tosteson and Tsiatis 1988). These methods do not apply for testing components of $\mathbf{x}$ which are measured precisely; see Carroll (1989) for examples. We can handle such problems using the techniques presented in Section 3.

Consider the hypothesis testing problem of the form

$$H_0\colon\ M\beta = 0 \quad \text{vs.} \quad H_1\colon\ M\beta \neq 0,$$

where $\mathbf{M}$ is a given r by p matrix of rank $r \leq p$. For instance, if we take $M = (1, 0, \ldots, 0)$, then $r = 1$ and we are testing whether or not the first variable in $\mathbf{x}$ affects the response $\mathbf{Y}$. Let $\hat{\beta}$ denote any estimator constructed in Section 3. To construct a Wald test for the hypothesis, we need a consistent estimate $\hat{\Sigma}(\hat{\beta})$ of $\Sigma(\hat{\beta})$, the asymptotic covariance of $\hat{\beta}$. For the least squares estimates of Sections 3.1, 3.3, 3.5, and 3.7, consistent estimation of $\Sigma(\hat{\beta})$ is easy: just substitute population quantities by their estimates. The only point that needs a little special care is that in going from the population versions to sample versions, terms of the form $\mathbf{w}_{i1} - \mathbf{w}_{i2}$ should be replaced by $\mathbf{w}_{i1} - \mathbf{w}_{i2} - m^{-1}\sum_{k=1}^{m}(\mathbf{w}_{k1} - \mathbf{w}_{k2})$.

For the SIR estimates of Sections 3.2, 3.4, 3.6, and 3.8, the same technique works once one estimates the vector $\mathbf{k}$

$= (k_1, \ldots, k_H)'$. To do this, multiply both sides of (2.7) by β', so that

$$c(y) = (\beta'\Sigma_{\mathbf{u}}\beta)^{-1}\beta'\{\zeta(y) - E(\mathbf{u})\},$$

from which it follows that $k_h = E\{c(\mathbf{Y})|\mathbf{Y} \in I_h\} = (\beta'\Sigma_{\mathbf{u}}\beta)^{-1}\beta'\{E\bar{\mathbf{u}}_h - E(\mathbf{u})\}$. Hence we can estimate k_h by

$$\hat{k}_h = (\hat{b}'_{\text{sir}}\hat{\Sigma}_{\mathbf{u}}\hat{b}_{\text{sir}})^{-1}\hat{b}'_{\text{sir}}(\bar{\hat{\mathbf{u}}}_h - \bar{\hat{\mathbf{u}}}),$$

where $\bar{\hat{\mathbf{u}}}_h$ denotes the hth slice mean for $\hat{\mathbf{u}}_i$'s. It is also easy to see that $\mathbf{k}'D\mathbf{k}$ can be estimated by $\hat{\mathbf{k}}'\hat{D}\hat{\mathbf{k}} = \hat{\lambda}(\hat{b}'_{\text{sir}}\hat{\Sigma}_{\mathbf{u}}\hat{b}_{\text{sir}})^{-1}$.

The Wald test at level α rejects the hypothesis if

$$\hat{\beta}'M'\{M\hat{\Sigma}(\hat{\beta})M'\}^{-1}M\hat{\beta} > \chi^2_r(1-\alpha),$$

where $\chi^2_r(1-\alpha)$ is the appropriate percentage point of the chi-squared random variable with r degrees of freedom.

5. GENERALIZATIONS

The method based on SIR is easy to generalize to other settings. First, instead of (1.1) we may consider a model

$$\mathbf{Y} = g(\alpha + \beta'\mathbf{x}, T, \varepsilon),$$

where T is a stratification variable such as sex or age group, so that g can be an arbitrary function with three arguments. We assume that the design distribution satisfies the linear conditional expectation condition (2.2) after conditioning on T; namely, for any direction b in R^p, there are functions of T, $c_0(T)$, $c_1(T)$, such that

$$E(b'\mathbf{x}|\beta'\mathbf{x}, T) = c_0(T) + c_1(T)\beta'\mathbf{x}.$$

It is clear that the inverse regression curves, $\eta(y|T) = E(\mathbf{w}|\mathbf{Y} = y, T)$ and $\zeta(y|T) = E(\mathbf{u}|\mathbf{Y} = y, T)$, still have the same property as given in Theorem 2.1 and (2.7):

$$\eta(y|T) = E(\mathbf{u}|T) + c(y|T)\Sigma_{\mathbf{u}|T}\beta, \qquad (2.7')$$

where $c(y|T) = (\beta'\Sigma_{\mathbf{x}|T}\beta)^{-1}E(\beta'(\mathbf{x} - E\mathbf{x})|\mathbf{Y} = y, T)$ and $\Sigma_{\mathbf{x}|T}$ is the conditional covariance of $\mathbf{x}$ given T. Thus, when creating slices, we need to use both $\mathbf{Y}$ and T. We can estimate β from each stratum of T and then combine the estimates.

Under the additional assumption that $\Sigma_{\mathbf{x}|T}$ does not depend on T, we can combine the estimate of the covariance Σ_ζ from each stratum:

$$\sum_t \sum_h \hat{p}_{h,t}(\bar{\mathbf{u}}_{h,t} - \bar{\mathbf{u}}_t)(\bar{\mathbf{u}}_{h,t} - \bar{\mathbf{u}}_t)',$$

where $\hat{p}_{h,t}$ is the proportion of the cases falling into slice h at stratum $T = t$, $\bar{\mathbf{u}}_{h,t}$ is the sample average of $\mathbf{u}$ for $T = t$ and slice h, and $\bar{\mathbf{u}}_t$ is the sample average of $\mathbf{u}$ for $T = t$. Then we can estimate β by the largest eigenvector of this matrix with respect to $\Sigma_{\mathbf{x}}$ as before. The asymptotic property will be similar to what was discussed earlier. The result of Hsing and Carroll (1992) can be used to justify the consistency property of the resulting estimate even if the number of cases per slice is small.

Another generalization is to consider the K component model (1.2). We can use the largest K eigenvectors of SIR (Step 3 of Section 3.2) to estimate the space spanned by the β's. We can justify our method based on the generalization of (2.7): $\eta(y) - E(\mathbf{u})$ falls into the space spanned by $\Sigma_{\mathbf{u}}\beta_1, \ldots, \Sigma_{\mathbf{u}}\beta_K$. The proof of this result follows along the same lines as the proof of Theorem 3.1 of Li (1991), incorporating the crucial property that $E(\delta|\mathbf{Y} = y, \mathbf{x}) = E(\delta) = 0$ as used in the proof of our Theorem 2.1.

6. DATA VISUALIZATION

Quite often a visual inspection of the data can lead to a suitable follow-up analysis, such as suggesting the functional form of g in (1.1), the detection of outliers, clustering analysis, and heterogeneity. When regressors are error free, one approach to visualizing regression data is based on regression diagnostics (Cook and Weisberg 1989). Alternatively, Li (1990a, 1990b, 1991) formed a framework for addressing this visualization issue, based on dimension reduction theory. He suggested SIR and several versions of principle Hessian direction (pHd) (Li 1990b) for accomplishing this goal.

When regressors are subject to error, even when $\mathbf{x}$ and $\mathbf{w}$ are scalar, plots of $\mathbf{Y}$ against $\mathbf{w}$ might give misleading information about the regression slope (or functions) of $\mathbf{Y}$ against $\mathbf{x}$ (Fuller 1987; Spiegelman 1986). This can be the case even if the measurement error is small (Carroll and Stefanski 1990). Despite these possibilities, in many instances curvature in $E(\mathbf{Y}|\mathbf{w})$ does reflect curvature in $E(\mathbf{Y}|\mathbf{x})$, and plotting $\mathbf{Y}$ against $\mathbf{w}$ may still be valuable. This section presents a theory for informatively visualizing the data when regressors are subject to error.

When $\mathbf{x}$ is observable, the best viewing angle is $\beta'\mathbf{x}$ under (1.1). But because $\mathbf{x}$ is not available, we can at best find a projection on $\mathbf{w}$ so that the projected variable has the highest correlation with $\beta'\mathbf{x}$ to ensure the best viewing angle obtainable from $\mathbf{w}$; that is, the one closest to the best view from $\mathbf{x}$. From Remark 2.4 at the end of Section 2, we see that $\beta'\mathbf{u}$, or its scalar multiple, is the desired variable. Because we can estimate the direction of β by $\hat{\beta}$, which denotes any estimate obtained earlier, we suggest plotting $\mathbf{Y}$ against $\hat{\beta}'\mathbf{u}$. If the correlation between $\beta'\mathbf{x}$ and $\beta'\mathbf{u}$ is high, then our plot may be useful—for instance, in suggesting the appropriate functional form in (1.1). When a validation sample of $(\mathbf{x}, \mathbf{w})$ is available, we can use it to estimate $\text{corr}(\beta'\mathbf{x}, \beta'\mathbf{u})$ by considering the sample correlation between $\hat{\beta}'\mathbf{x}$ and $\hat{\beta}'\mathbf{u}$.

The following is a simulation example to see how effective our method is. We consider two models for generating the data:

$$\mathbf{Y} = (\alpha + \beta'\mathbf{x} + \varepsilon)^2 \qquad (6.1)$$

and

$$\mathbf{Y} = (\alpha + \beta'\mathbf{x})^2 + \varepsilon. \qquad (6.2)$$

The first model falls into the Box–Cox transformation family; for the second model the transformation is taken only for the mean response function, a special case of a "transform-both-sides" model considered by Carroll and Ruppert (1988; cf. our Remark 6.1). The coordinates of $\mathbf{x}$ and ε are independent standard normal random variables. We set the dimension parameters as $p = 6 = q$, the primary sample size

Table 1. Summary of $\hat{b}_{sir} = (\hat{b}_1, \ldots, \hat{b}_6)'$ for Model (6.1)

Number of slices	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	cos(x)	corr(w)
5	.571	.567	.569	.001	−.002	.007	.986	.986
	(.069)	(.059)	(.059)	(.075)	(.071)	(.079)	[.949]	[.950]
10	.571	.568	.570	−.002	−.005	.004	.987	.988
	(.067)	(.052)	(.058)	(.069)	(.070)	(.077)	[.959]	[.959]
20	.571	.569	.568	.005	−.005	.006	.986	.987
	(.066)	(.054)	(.062)	(.073)	(.067)	(.078)	[.950]	[.949]

NOTE: Standard deviation and minimum are in parentheses and brackets.

as $n = 200$, the validation sample size as $m = 100$, and $\beta = (1, 1, 1, 0, 0, 0)'$, $\alpha = 4$. The relationship between $\mathbf{w}$ and $\mathbf{x}$ will be governed by the linear model (2.1), with $p = q = 6$, $\Gamma = I$. The distribution of δ is normal with mean 0 and covariance being a diagonal matrix with diagonal element $(0, 1/3, 1/3, 1/3, 1/3, 1/3)$. Thus each coordinate of $\mathbf{x}$, except for the first one, is contaminated by an error of a size equal to .577 of its standard deviation.

Because our procedure does not require knowledge of the functional form that generates the data, we can apply it to both (6.1) and (6.2). After 100 simulation runs, we summarize the results for $\hat{b}_{\text{sir}}$ in Tables 1 and 2. Because we are interested only in estimating the direction of β, we have standardized our estimate to have unit length. For each model the mean of $\hat{b}_{\text{sir}}$ is very close to the theoretical value $(.577, .577, .577, 0, 0, 0)' = \beta/\|\beta\|$. This demonstrates that our procedure can avoid the bias that one normally would expect to see without proper model specification. For instance, the naive estimate of regressing the square root of $\mathbf{Y}$ against $\mathbf{w}$ for data generated by (6.1) gives the slope $(1, .75, .75, 0, 0. 0)'$ on average, which is not proportional to β. The standard deviation of our estimate is reasonably small compared, for instance, to the ideal value of $.07 \approx 1/\sqrt{n}$, the standard deviation of the least squares estimate for β under (6.1) if the square transformation were known and if $\mathbf{x}$ were observed without errors. The cosine of the angle between $\hat{b}_{\text{sir}}$ and β, which is the same as the correlation coefficient between $\hat{b}'_{\text{sir}}\mathbf{x}$ and $\beta'\mathbf{x}$, is very close to 1, with the lowest value in the neighborhood of .95; see the next to the last column in each table. We considered three choices of the number of slices $H = 5$, 10, 20, with an equal number of observation per slice. This illustrates the stability of our procedure in regard to the change in H at least for this example.

Now to demonstrate the application of our method to data visualization, a single additional run is taken. For this realization, we found that $\hat{b}_{\text{sir}} = (-.54, -.64, -.53, -.03, -.09, -.01)'$ for model (6.1) and $(-.57, -.58, -.56, .04, -.03, -.00)'$ for model (6.2). The estimate of the transformation L is as follows:

$$\begin{pmatrix} 1.00 & .000 & .000 & .000 & .000 & .000 \\ .077 & .765 & .030 & .104 & .012 & .023 \\ .038 & -.006 & .785 & .013 & .013 & -.020 \\ .012 & .015 & .048 & .760 & -.048 & .107 \\ .059 & -.0619 & -.044 & -.063 & .704 & -.047 \\ .037 & -.010 & -.059 & .065 & .046 & .747 \end{pmatrix}$$

Then we compute $\hat{b}'_{\text{sir}}\hat{\mathbf{u}}_i = (\hat{L}\hat{b}_{\text{sir}})'\mathbf{w}_i$ for $i = 1, \ldots, n$. The scatterplots of $\mathbf{Y}_i$ against $\hat{b}'_{\text{sir}}\hat{\mathbf{u}}_i$ are given in Figures 1 and 4, for models (6.1) and (6.2). These plots suggest that analysis based on transformation is reasonable to pursue further.

In Figures 2 and 5 we plot $\mathbf{Y}_i$ against $\beta'\mathbf{u}_i$, the best view on $\mathbf{Y}$ from the surrogate variable if L and β were known. Compared to what we had seen in Figures 1 and 4, we see that very little has been lost due to the estimation. The correlation between the variable $\beta'\mathbf{u}$ and our estimated variable $\hat{\beta}'\mathbf{u}$ is as high as .99. The last column in each of Tables 1 and 2 gives the mean and the lowest possible value of this correlation over the same 100 simulation runs done earlier. The lowest number is still as high as .95. This suggests that approximately the same view would be obtained from other simulation runs.

We also provide plots of $\mathbf{Y}_i$ against $\beta'\mathbf{x}_i$ in Figures 3 and 6, the best view of the models from the uncontaminated regressor. We find that the views obtained by our estimate, Figures 1 and 4, are also very close to these best views. This can be attributed to the fact that for our parameter setting, despite the apparent heavy contamination rate, the correlation between $\beta'\mathbf{x}$ and $\beta'\mathbf{u}$ is high, equaling to about .91. Figures 3 and 6 are supposed to be close to Figures 2 and 5,

Table 2. Summary of $\hat{b}_{sir} = (\hat{b}_1, \ldots, \hat{b}_6)'$ for Model (6.2)

Number of slices	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	cos(x)	corr(w)
5	.569	.573	.571	.007	.000	.002	.989	.990
	(.059)	(.047)	(.053)	(.060)	(.072)	(.067)	[.968]	[.971]
10	.573	.571	.572	.005	−.001	.004	.990	.991
	(.060)	(.045)	(.051)	(.057)	(.065)	(.062)	[.974]	[.972]
20	.570	.573	.573	.006	−.004	.003	.990	.991
	(.059)	(.045)	(.053)	(.057)	(.064)	(.061)	[.970]	[.968]

NOTE: Standard deviation and minimum are in parentheses and brackets.

Figure 1. SIR's View for Model (6.1).

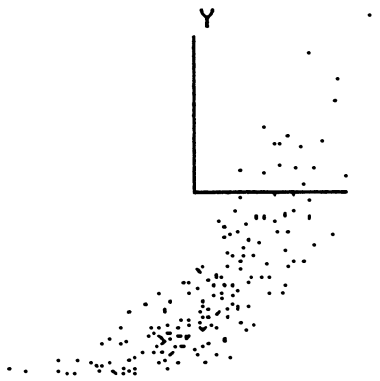

Figure 3. Best View on (6.1) From x.

which in turn have been shown to be similar to Figures 1 and 4.

If we have a small number of additional data points available on $(\mathbf{Y}, \mathbf{x})$, then we can plot $\mathbf{Y}$ against $\hat{\beta}'\mathbf{x}$ as well. To illustrate this, we generate 10 new data points for $(\mathbf{Y}, \mathbf{x})$ from (6.2). The plot is given in Figure 7, which shows a quadratic trend well.

Remark 6.1. Carroll and Ruppert (1988) considered the transform-both-sides model, which allows another transformation on $\mathbf{Y}$ before getting a model like (6.2). Because our procedure is invariant under the monotone transformation, we would have obtained the same estimate if our data had been given after applying any unknown monotone transformation to $\mathbf{Y}$ in model (6.2).

Remark 6.2. When studying the plot of $\mathbf{Y}_i$ against $\hat{b}'_{\text{sir}}\hat{\mathbf{u}}_i$, it is worthwhile to pay some attention to the inverse regression curve; namely, smoothing $\hat{b}'_{\text{sir}}\hat{\mathbf{u}}_i$ against $\mathbf{Y}_i$. This is because from Theorem 2.1, the curve $E(\beta'\mathbf{u}|\mathbf{Y} = y)$ is an affine transformation of the curve $E(\beta'\mathbf{x}|\mathbf{Y} = y)$. This curve is particularly helpful if the data are indeed generated from a model $\mathbf{Y} = g(\beta'\mathbf{x}) + \varepsilon$, with the standard deviation of ε being small enough so that we can approximate $E(\beta'\mathbf{x}|\mathbf{Y} = y)$ by $g^{-1}(y)$. The smoothed inverse regression curve is expected to be close to the inverse of g. We can use the inverse of the smoothed inverse regression curve to suggest a suitable functional form for g.

Remark 6.3. If α were set at 0, then we cannot estimate the direction of β well because the theoretical inverse regression curve will degenerate to a point. Similar to the regressor-error-free case, there are several variants of second-moment-based SIR or pHd methods (Li 1990b) available for remedy.

7. THE CONSTANT OF PROPORTIONALITY

As mentioned previously, when g is completely unknown we can at most identify β up to a constant of proportionality. We could use the data visualization technique as discussed in Section 6 for suggesting a reasonable functional form. On the other hand, if g is given, then we shall discuss how to estimate the entire β based on the reduced data.

Let $\tilde{\beta}$ be the vector to which any of our estimates $\hat{\beta}$ constructed in Section 3 converges. We have shown that $\gamma\tilde{\beta} = \beta$ for some constant γ. Let $\mathbf{z} = \tilde{\beta}'\mathbf{x}$. Then we can write (1.1) as $\mathbf{Y} = g(\alpha + \gamma\mathbf{z}, \varepsilon)$. Our job now is to estimate (α, γ). Let $\tau = \tilde{\beta}'\mathbf{u}$. As discussed in Section 6, τ is in a sense most informative in predicting $\mathbf{z}$. We can consider τ to be the surrogate for $\mathbf{z}$. If $\tilde{\beta}$ were estimated without error, then based on the primary sample $(\mathbf{Y}_i, \tau_i)$, $i = 1, \ldots, n$ and the validation sample $(\mathbf{z}_i, \tau_i)$, $i = n + 1, \ldots, n + m$, where $\tau_i = \tilde{\beta}'u_i$ and $\mathbf{z}_i = \tilde{\beta}'\mathbf{x}_i$, we could apply many available techniques, such as the parametric method of Carroll et al. (1984), the semiparametric method of Stefanski and Carroll (1987), the small measurement error asymptotics of Stefanski and Carroll (1990), and the nonparametric kernel regression method of Carroll and Wand (1990). We suggest that after estimating $\tilde{\beta}$ by $\hat{\beta}$, we construct $\hat{\tau}_i = \hat{\beta}'\hat{\mathbf{u}}_i$ and $\hat{\mathbf{z}}_i = \hat{\beta}'\mathbf{x}_i$ to replace τ_i, $\mathbf{z}_i$ and apply any of these methods. Further work on this suggestion is necessary.

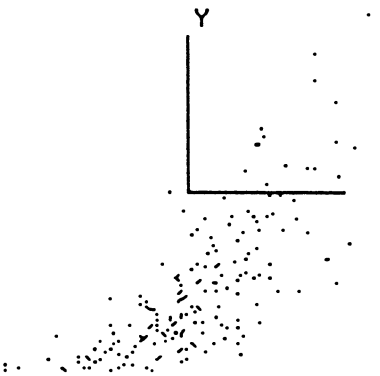

Figure 2. Best View for (6.1) From w.

Figure 4. SIR's View for Model (6.2).

Figure 5. Best View for (6.2) From w.

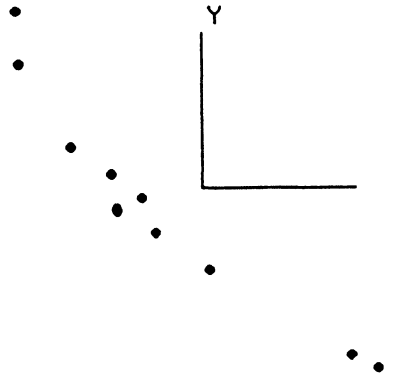

Figure 7. SIR's View for (6.2) From New (y, x).

8. EXAMPLE

We have access to a restricted data set containing breast cancer incidence **Y** and **x** = (age, body mass, nutrient intake), the latter in this case being the logarithm of saturated fat. The primary data set was of size $n = 2{,}800$, and the validation data were of size $m = 650$. The fallible version of nutrient intake was assessed in the study by an interview detailing the previous day's diet; the version of truth used here was the average of three such interviews. The measurement error is quite large, with fully more than 50% of the observed variability in fat in error. All measurement error analyses of these data performed previously have shown a large age effect and a negligible body mass effect. Some analyses show a significant effect due to the nutrient intake, and others do not; in all cases, the coefficient has been negative. For purposes of this numerical illustration, we will assume representative validation. Full details will be provided elsewhere. If for no other reason that there were only 59 reported cases of breast cancer in this study, the results should be treated with extreme caution.

Programming the methods discussed in this article is very easy. This is particularly the case for binary regression, because the least squares and SIR methods discussed in Section 3 yield identical estimates of β, in terms of unit length. In this example the ordinary logistic regression estimate of unit length obtained from regressing **Y** on **w** was (.97, −.12, −.20), with two-sided significance levels (.00, .68, .06). Our methods yielded estimates (.87, −.14, −.47), with two-sided significance levels (.00, .64, .02). Note that this measurement error analysis yielded a larger estimated relative effect due to the nutrient than did the ordinary logistic regression. The difference in statistical significance for the saturated fat coefficient may be due to the use of information standard errors for the ordinary logistic analysis; an analysis using M estimator techniques to construct standard errors yielded lower significance levels. For example, if we assume that given observed fat, the true fat is independent of age and body mass (checked by a linear regression analysis), then the method of Rosner et al. (1989) applied to these data yielded estimates (.91, −.11, −.40) and significance levels (.00, .64, .01).

Figure 6. Best View for (6.2) From x.

APPENDIX A: PROOF OF FISHER CONSISTENCY IN CONVEX REGRESSION, REMARK 2.6

We use the conditional argument similar to that for the proof of Theorem 2.1 in Li and Duan (1989). For any $b \in R^p$, by Jensen's inequality we shall have

$$\begin{aligned} E\rho(\mathbf{Y}, a + b'\mathbf{u}) &= E[E\{\rho(\mathbf{Y}, a + b'\mathbf{u}) | \beta'\mathbf{x}, \varepsilon, \beta'\mathbf{u}\}] \\ &\geq E[\rho\{\mathbf{Y}, a + E(b'\mathbf{u} | \beta'\mathbf{x}, \beta'\mathbf{u})\}] \\ &= E[\rho\{\mathbf{Y}, a + E(b'\mathbf{u} | \beta'\mathbf{u})\}] \\ &= E\{\rho(\mathbf{Y}, a + c_0 + c_1\beta'\mathbf{u})\}, \\ &\quad \text{for some real numbers } c_0, c_1. \end{aligned}$$

Here the second to last equality is due to the fact that given $\beta'\mathbf{u}$, **u** is independent of $\beta'\mathbf{x}$, a consequence of the joint normality of **x** and **w**. It is now clear that a minimizer can be found along the direction of β, proving our claim.

APPENDIX B: APPROXIMATION OF (3.5) BY (3.6)

Recall that $\Delta_{\mathbf{u}} = \bar{\mathbf{U}} - E\bar{\mathbf{U}}$. The term inside the parentheses of (3.5) can be written as

$$\bar{\mathbf{U}}\hat{D}\mathbf{k} = \sum_{h=1}^{H} \hat{p}_h k_h \bar{\mathbf{u}}_h = n^{-1} \sum_{i=1}^{n} \tilde{\mathbf{Y}}_i \mathbf{u}_i.$$

Comparing this with (3.6), it remains to show that $(n^{-1}\sum_{i=1}^{n} \tilde{\mathbf{Y}}_i)\bar{\mathbf{u}}$ is of the order $O_p(n^{-1})$. But this follows directly from our assumption that $E\mathbf{u} = LE\mathbf{w} = 0$ and the fact that

$$E\tilde{\mathbf{Y}}_i = \sum_{h=1}^{H} p_h k_h = Ec(\mathbf{Y}) = 0,$$

where the last equality follows from Theorem 2.1.

APPENDIX C: PROOF OF LEMMA 3.1

Let $\mathbf{z}_i = \mu + e_i$. Then it follows that

$$\frac{\mathbf{z}_i}{\|\mathbf{z}_i\|} - \frac{\mu}{\|\mu\|} = \frac{e_i}{\|\mu\|} - \frac{\mu\langle\mu, \varepsilon\rangle}{\|\mu\|^3} + o_p(n^{-1/2}) = \frac{P_2 e_i}{\|\mu\|} + o_p(n^{-1/2}).$$

It is now easy to see that Lemma 3.1 holds.

APPENDIX D: PROOF OF (3.20)

Define

$$A = (\hat{\Omega}^{-1} - \Omega^{-1})\Gamma'(\Sigma_w - \Sigma_\delta)\Gamma\Omega^{-1}$$
$$= -\Omega^{-1}(\hat{\Omega} - \Omega)\Sigma_x + O_p(n^{-1}),$$
$$B = \Omega^{-1}(\hat{\Gamma} - \Gamma)'(\Sigma_w - \Sigma_\delta)\Gamma\Omega^{-1} = \Omega^{-1}(\hat{\Gamma} - \Gamma)'\Gamma\Sigma_x,$$

and

$$A + B = -\Omega^{-1}\Gamma'(\hat{\Gamma} - \Gamma)\Sigma_x + O_p(n^{-1}).$$

Then

$$\hat{\Sigma}_x - \Sigma_x = (A + B) + (A + B)' + \Omega^{-1}\Gamma'(\hat{\Sigma}_{1w} - \Sigma_w)\Gamma\Omega^{-1} - \Omega^{-1}\Gamma'(\hat{\Sigma}_\delta - \Sigma_\delta)\Gamma\Omega^{-1}.$$

Hence

$$\hat{L} - L = [\{(A + B) + (A + B)'\}\Gamma' + \Omega^{-1}\Gamma'(\hat{\Sigma}_{1w} - \Sigma_w)\Gamma\Omega^{-1}\Gamma' - \Omega^{-1}\Gamma'(\hat{\Sigma}_\delta - \Sigma_\delta)\Gamma\Omega^{-1}\Gamma' + \Sigma_x(\hat{\Gamma} - \Gamma)' - \Sigma_x\Gamma'\Sigma_w^{-1}(\hat{\Sigma}_{1w} - \Sigma_w)]\Sigma_w^{-1}$$
$$= n^{-1}\sum_{i=1}^{n} l_i + m^{-1}\sum_{i=n+1}^{n+m} \tau_i + O_p(n^{-1} + m^{-1}).$$

The proof now follows from (3.10).

[*Received October 1990. Revised December 1991.*]

REFERENCES

Brillinger, D. R. (1977), "The Identification of a Particular Nonlinear Time Series System," *Biometrika,* 64, 509–515.

——— (1983), "A Generalized Linear Model With Gaussian Regressor Variables," in *A Festchrift for Erick L. Lehmann,* Belmont, CA: Wadsworth, pp. 97–114.

Carroll, R. J. (1989), "Covariance Analysis in Generalized Linear Measurement Error Models," *Statistics in Medicine,* 8, 1075–1093.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression,* London: Chapman and Hall.

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

Carroll, R. J., and Stefanski, L. A. (1990), "Approximate Quasi-likelihood Estimation in Models With Surrogate Predictors," *Journal of the American Statistical Association,* 85, 652–663.

Carroll, R. J., and Wand, M. P. (1991), "Semiparametric Estimation in Logistic Measurement Error Models," *Journal of the Royal Statistical Association,* Ser. B, 53, 573–585.

Chen, H. (1991), "Estimation of a Projection Pursuit Type Regression Model," *The Annals of Statistics,* 19, 142–157.

Cook, R. D., and Weisberg, S. (1989), "Regression Diagnostics With Dynamic Graphics" (with discussion), *Technometrics,* 31, 277–308.

Diaconis, P., and Freedman, D. (1984), "Asymptotics of Graphical Projection Pursuit," *The Annals of Statistics,* 12, 793–815.

Duan, N., and Li, K. C. (1991), "Slicing Regression: A Link-Free Regression Method," *The Annals of Statistics,* 19, 505–530.

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

Hall, P. (1989), "On Projection Pursuit Regression," *The Annals of Statistics,* 17, 573–588.

Hall, P., and Li, K. C. (1993), "On Almost Linearity of Low-Dimensional Projections from High-Dimensional Data," unpublished manuscript submitted to *Annals of Statistics.*

Härdle, W., and Stoker, T. M. (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association,* 84, 986–995.

Hsing, T., and Carroll, R. J. (1992), "Asymptotic Properties of Sliced Inverse Regression," *The Annals of Statistics,* 20, 1040–1061.

Li, K. C. (1990a), "Data Visualization With SIR: A Transformation Based Projection Pursuit Method," *UCLA Statistical Series #24.*

——— (1992), "On Principal Hessian Direction for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association,* in press.

——— (1991), "Sliced Inverse Regression for Dimension Reduction" (with discussion), *Journal of the American Statistical Association,* 86, 316–342.

Li, K. C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics,* 17, 1009–1052.

Pepe, M. S., and Fleming, T. R. (1991), "A Nonparametric Method for Dealing with Mismeasured Covariate Data," *Journal of the American Statistical Association,* 86, 108–113.

Pierce, D. A., Stram, D. O., Vaeth, M., and Schafer, D. W. (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, 351–359.

Rosner, B., Willett, W. C., and Spiegelman, D. (1989), "Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error," *Statistics in Medicine,* 8, 1051–1070.

Spiegelman, C. H. (1986), "Two Pitfalls of Using Standard Regression Diagnostics When Both X and Y Have Measurement Error," *The American Statistician,* 40, 245–248.

Stefanski, L. A., and Carroll, R. J. (1990), "Score Tests in Generalized Linear Measurement Error Models," *Journal of the Royal Statistical Society,* Ser. B, 52, 345–360.

Tosteson, T., Stefanski, L. A., and Schafer, D. W. (1989), "A Measurement Error Model for Binary and Ordinal Regression," *Statistics in Medicine,* 8, 1139–1147.

Tosteson, T., and Tsiatis, A. (1988), "The Asymptotic Relative Efficiency of Score Tests in a Generalized Linear Model With Surrogate Covariates," *Biometrika,* 75, 507–514.