THE IMPORTANCE OF ASSESSING MEASUREMENT RELIABILITY IN MULTIVARIATE REGRESSION

by

Leon Jay Gleser
Purdue University

Technical Report #88-19
Preliminary Report

Department of Statistics
Purdue University

May 1988

*

THE IMPORTANCE OF ASSESSING MEASUREMENT RELIABILITY IN MULTIVARIATE REGRESSION

LEON JAY GLESER *

In many contexts in which multivariate linear regression models are applied, some or all of the independent (predictor) variables are measured with error, and the goal is to assess the relationship of the dependent variables to the true predictor variables. When this is the case, it is argued that it is extremely important to determine the reliability matrix Λ of the measurement X of the vector of predictors. If Λ is singular, the slope matrix B is not identifiable. If Λ is nonsingular, but nearly singular, B cannot be accurately estimated. Using Λ, or an estimate of Λ, one can construct an estimator of B which is a simple adjustment of the classical least squares estimator of slope. It is shown that some of the unpleasant behavior (nonexistence of expected values) of estimators of B stem from similar behavior of estimates used for Λ. Also discussed is a useful canonical reliability analysis of the predictor variables which closely resembles the principal component pre–analyses of predictors which are advocated for spotting and treating multicollinearity problems in classical linear regression.

KEY WORDS: Errors–in–variables, linear structural relationships, correction for attenuation.

* Leon Jay Gleser is Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907. This research was supported in part by National Science Foundation Grant DMS 8501966.

1. INTRODUCTION

In many applications of multivariate linear regression, some (or all) of the components of the r–dimensional (row) vector x of predictors are known to be measured with error. If the goal of investigation is to assess the relationship of the p–dimensional vector Y of dependent variables to the true predictor x, rather than to the observed predictor X, then the classical least squares method produces biased estimates. (Cochran, 1968; Fuller, 1987).

Alternative methods of inference are based on multivariate linear *errors–in–variables regression* (EIVR) models of the form:

$$Y_i = \alpha + x_i B + e_i, \quad X_i = x_i + f_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where the subscript i indexes the observation, f_i is the vector of measurement errors for the true predictor x_i, and Y_i, X_i are the observed vectors of dependent variables and predictor variables, respectively. The vector e_i is a residual error vector for the regression of Y_i on x_i; it may include both errors of measurement for a true (latent) vector y_i of dependent variables, and residuals from the regression of y_i on x_i (called *equation errors* in Fuller, 1987).

It is usually assumed that for given $x_1, \ldots, x_n$, the vectors $\varepsilon_i = (e_i, f_i)$ of random errors are (conditionally) independent and identically distributed (i.i.d.) with common mean vector $\mathbf{0}$ and common covariance matrix

$$\Sigma_{\varepsilon\varepsilon} = \begin{pmatrix} \Sigma_{ee} & \Sigma_{ef} \\ \Sigma_{fe} & \Sigma_{ff} \end{pmatrix}.$$

The vectors $x_1, x_2, \ldots, x_n$ of true predictors can be regarded either as fixed unknown parameters (*functional case*), or as i.i.d. random vectors with common mean vector μ and covariance matrix Σ_{xx} (*structural case*).

When all random vectors (ε_i, and possibly x_i) in the model have multivariate normal distributions, the parameters $\alpha, \mu, B, \Sigma_{xx}, \Sigma_{\varepsilon\varepsilon}$ of the model are not identifiable without restrictions. Approaches to estimation must take this fact into account. That is, *the normal* *distributions in linear EIVR models play a "worst-case" role in robustness considerations.* For this reason, we will discuss estimation in the model (1.1) under multivariate normality assumptions on the error vectors ε_i. Further, we will concentrate on the structural case of the EIVR model (1.1), and assume that the x_i's are normally distributed. Under reasonable assumptions on the sequence $\{x_1, x_2, \ldots\}$ of true predictor vectors in functional linear EIVR models, maximum likelihood estimators derived from the corresponding *structural* EIVR model are asymptotically efficient in the functional case context as $n \to \infty$ (Gleser, 1983; Nussbaum, 1984; Bickel and Ritov, 1987). Hence, study of estimation in the structural case has implications for estimation in the functional case.

To summarize, it is assumed that:

$$\begin{aligned}
&\varepsilon_i = (e_i, f_i) \text{ are i.i.d. MVN } (0, \Sigma_{\varepsilon\varepsilon}),\\
&x_i \text{ are i.i.d. MVN } (\mu, \Sigma_{xx}),\\
&\varepsilon_1, \ldots, \varepsilon_n, x_1, \ldots, x_n \text{ are mutually independent.}
\end{aligned} \tag{1.2}$$

The parameters of the model are $\alpha, \mu, B, \Sigma_{xx}$ and $\Sigma_{\varepsilon\varepsilon}$. Our primary interest is in estimation of the $r \times p$ matrix B of slopes.

Identifiability

Let

$$Z_i = (Y_i, X_i), \quad i = 1, 2, \ldots, n.$$

Under the assumptions (1.1), (1.2), the Z_i are i.i.d. MVN (η, Ψ), where

$$\eta = (\alpha + \mu B, \mu), \quad \Psi = \begin{pmatrix} B'\Sigma_{xx}B + \Sigma_{ee} & B'\Sigma_{xx} + \Sigma_{ef} \\ \Sigma_{xx}B + \Sigma_{fe} & \Sigma_{xx} + \Sigma_{ff} \end{pmatrix}.$$

Note that there are $\frac{1}{2}(p+r)(p+r+3)$ free parameters in η, Ψ, and that these parameters determine the joint distribution of the observed vectors $Z_1, \ldots, Z_n$. On the other hand, there are a total of

$$\frac{1}{2}(p+r)(p+r+3) + rp + \frac{1}{2}r(r+1)$$

free parameters in the parameters $\alpha, \mu, B, \Sigma_{xx}, \Sigma_{\varepsilon\varepsilon}$ of the model (1.1), (1.2). Thus, the model (1.1), (1.2) is overparameterized, and a minimum of $rp + \frac{1}{2}r(r+1)$ restrictions are needed to parametrically identify the model.

It is not difficult to see that B is confounded with the slope matrix $\Theta = \Sigma_{ff}^{-1}\Sigma_{fe}$ of the regression of the error vectors e_i on the error vectors f_i. Indeed, we can represent the observations Z_i in two ways:

$$Z_i = (\alpha, \mathbf{0}) + x_i(B, I_r) + f_i(\Theta, I_r) + (e_i - f_i\Theta, \mathbf{0}),$$

and

$$Z_i = (\alpha + \mu(B - \Theta), \mathbf{0}) + (f_i + \mu)(\Theta, I_r) + (x_i - \mu)(B, I_r) + (e_i - f_i\Theta, \mathbf{0}),$$

where I_r is the r–dimensional identity matrix. In the second of these two representations, the $(f_i + \mu)$ play the role of the true predictors x_i and Θ plays the role of B. Since x_i, f_i and $e_i - f_i\Theta$ are mutually independent, both of the above representations yield the same common distribution for the Z_i's.

Barring any way of telling x_i from f_i, it appears necessary to know the value of Θ. Fortunately, in many applications Y_i and X_i are separately measured, and it is reasonable to assume that

$$\Sigma_{fe} = \mathbf{0}. \tag{1.3}$$

In this case, $\Theta = \mathbf{0}$. Note that (1.3) imposes pr restrictions on the parameters of the model (1.1), (1.2), leaving $\frac{1}{2}r(r+1)$ further restrictions to be determined.

Note. If Θ is known and is not equal to zero, we can reduce to the case (1.3) by the transformation $Y_i \to \tilde{Y}_i = Y_i - X_i\Theta$. This yields the representation

$$\tilde{Y}_i = \alpha + x_i\tilde{B} + \tilde{e}_i, \quad X_i = x_i + f_i,$$

where $\tilde{B} = B - \Theta, \tilde{e}_i = e_i - f_i\Theta$. Note that cov $(\tilde{e}_i, f_i) = \mathbf{0}$, and that the covariance matrix of $\tilde{e}_i$ is $\Sigma_{\tilde{e}\tilde{e}} = \Sigma_{ee} - \Theta'\Sigma_{ff}\Theta$. Estimates of $\tilde{B}, \Sigma_{\tilde{e}\tilde{e}}$ and Σ_{xx} obtained from $\tilde{Y}_i, X_i$ directly yield estimates of $B = \tilde{B} + \Theta$ and $\Sigma_{ee} = \Sigma_{\tilde{e}\tilde{e}} + \Theta'\Sigma_{xx}\Theta$.

From dimensionality considerations, it is natural to impose the remaining $\frac{1}{2}r(r+1)$ restrictions on the parameters of the common distribution of the r–dimensional observations X_i. If this approach is followed, it is shown in Section 2 that the *reliability matrix*

$$\Lambda = \Sigma_{XX}^{-1}\Sigma_{xx} = (\Sigma_{xx} + \Sigma_{ff})^{-1}\Sigma_{xx} \tag{1.4}$$

plays an important role in the estimation of B. Indeed, regardless of the particular restrictions used (as long as they only restrict the parameters of the marginal distribution of the X_i's), the maximum likelihood estimator $\hat{B}_{\text{MLE}}$ of B has the form

$$\hat{B}_{\text{MLE}} = \hat{\Lambda}^{-1}\hat{B}_{\text{LSE}}, \tag{1.5}$$

where $\hat{B}_{\text{LSE}}$ is the classical least squares estimator of the slope matrix of the regression of Y_i on X_i, $i = 1, 2, \ldots, n$, and $\hat{\Lambda}$ is the maximum likelihood estimator of Λ (based only on the X_i's). It is further shown that the expected value, $E(\hat{B}_{\text{MLE}})$, of $\hat{B}_{\text{MLE}}$ exists if and only if the expected value of $\hat{\Lambda}^{-1}$ exists. This result provides an explanation for the nonexistence of integral moments of $\hat{B}_{\text{MLE}}$ when, as is often the case in EIVR contexts, this unpleasant property occurs; and further suggests adjustment of $\hat{\Lambda}^{-1}$ as a way to obtain better–behaved estimators of B.

In Section 3, it is shown that the reliability matrix Λ contains key information about the accuracy with which one can estimate the slope matrix B. For example, the characteristic roots and vectors of Λ serve to identify linear combinations of B which can be most (or least) accurately estimated.

The results in Sections 2 and 3, indicate the importance of assessing measurement reliability (in terms of the reliability matrix Λ) in applications of multivariate linear regression.

2. MAXIMUM LIKELIHOOD ESTIMATION

It follows from (1.1), (1.2) and (1.3) that the vectors $Z_i = (Y_i, X_i)$, $i = 1, 2, \ldots, n$, of observations are i.i.d. with a common $(p + r)$–dimensional normal distribution having mean vector $\eta = (\alpha + \mu B, \mu)$ and covariance matrix

$$\Psi = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} = \begin{pmatrix} B'\Sigma_{xx}B + \Sigma_{ee} & B'\Sigma_{xx} \\ \Sigma_{xx}B & \Sigma_{xx} + \Sigma_{ff} \end{pmatrix}.$$

Consequently, the conditional distribution of Y_i given X_i is multivariate normal with mean vector

$$E[Y_i|X_i] = \alpha + \mu B + (X_i - \mu)(\Sigma_{xx} + \Sigma_{ff})^{-1}\Sigma_{xx}B$$

and covariance matrix

$$\Sigma_{YY\cdot X} = \Sigma_{ee} + B'\Sigma_{xx}B - B'\Sigma_{xx}(\Sigma_{xx} + \Sigma_{ff})^{-1}\Sigma_{xx}B.$$

From (1.4), $\Lambda = (\Sigma_{xx} + \Sigma_{ff})^{-1}\Sigma_{xx}$. Thus

$$E[Y_i|X_i] = \xi + X_i\Lambda B, \tag{2.1}$$

where

$$\xi = \alpha + \mu(I_r - \Lambda)B.$$

It follows from (2.1) that the classical least squares estimator $\hat{B}_{\text{LSE}}$ of the slope matrix for the regression of Y_i on X_i is an unbiased estimator of

$$\boldsymbol{\beta} = \Lambda B, \tag{2.2}$$

rather than of B, and hence is biased unless $\Lambda = I_r$.

We have already implicitly assumed that the covariance matrix $\Sigma_{XX} = \Sigma_{xx} + \Sigma_{ff}$ of X_i is nonsingular. Let us also assume that the common covariance matrix Σ_{xx} of the true predictors x_i is nonsingular, so that Λ is invertible. Consider the reparameterization

$$(\alpha, \mu, B, \Sigma_{xx}, \Sigma_{ee}, \Sigma_{ff}) \to (\xi, \mu, \boldsymbol{\beta}, \Sigma_{xx}, \Sigma_{YY\cdot X}, \Sigma_{ff}). \tag{2.3}$$

This reparameterization is one–to–one onto since

$$\begin{aligned} \alpha = \xi - \mu\Sigma_{xx}^{-1}\Sigma_{ff}\boldsymbol{\beta}, \quad B = \Sigma_{xx}^{-1}(\Sigma_{xx} + \Sigma_{ff})\boldsymbol{\beta} \\ \Sigma_{ee} = \Sigma_{YY\cdot X} - \boldsymbol{\beta}'\Sigma_{ff}\Sigma_{xx}^{-1}(\Sigma_{xx} + \Sigma_{ff})\boldsymbol{\beta}. \end{aligned} \tag{2.4}$$

We can write the density of Z_i in terms of the new parameterization as

$$\begin{aligned} &f(Z_i|\xi, \mu, \boldsymbol{\beta}, \Sigma_{xx}, \Sigma_{YY\cdot X}, \Sigma_{ff}) \\ &\quad = f(Y_i|X_i, \xi, \boldsymbol{\beta}, \Sigma_{YY\cdot X})f(X_i|\mu, \Sigma_{xx}, \Sigma_{ff}). \end{aligned} \tag{2.5}$$

Let

$$\overline{Z} = (\overline{Y}, \overline{X}) = n^{-1}\sum_{i=1}^{n}(Y_i, X_i) = n^{-1}\sum_{i=1}^{n} Z_i,$$

and

$$W = \begin{pmatrix} W_{YY} & W_{YX} \\ W_{XY} & W_{XX} \end{pmatrix} = \sum_{i=1}^{n} (Z_i - \overline{Z})'(Z_i - \overline{Z}).$$

Theorem 2.1. For any parametric restrictions placed on $(\Sigma_{xx}, \Sigma_{ff})$ sufficient to identify these parameters, the maximum likelihood estimators $\hat{\Sigma}_{xx}, \hat{\Sigma}_{ff}$ of Σ_{xx}, Σ_{ff} are a function of the data only through W_{XX}. Further, the maximum likelihood estimator of μ is $\hat{\mu} = \overline{X}$. When $\hat{\Sigma}_{xx}$ is nonsingular, the maximum likelihood estimator of B is

$$\hat{B}_{\text{MLE}} = \hat{\Lambda}^{-1}\hat{B}_{\text{LSE}} = \hat{\Lambda}^{-1} W_{XX}^{-1} W_{XY}, \tag{2.6}$$

where $\hat{\Lambda} = (\hat{\Sigma}_{xx} + \hat{\Sigma}_{ff})^{-1}\hat{\Sigma}_{xx}$. Finally, maximum likelihood estimators of the remaining parameters are

$$\hat{\alpha} = \overline{Y} - \overline{X}[I_r + (\hat{\Lambda}')^{-1}(I_r - \hat{\Lambda})]\hat{B}_{\text{LSE}}$$

and

$$\hat{\Sigma}_{ee} = \frac{1}{n}\left[W_{YY} - W_{YX}W_{XX}^{-1}W_{XY}\right] - \hat{B}'_{\text{LSE}}\hat{\Sigma}_{ff}\hat{\Sigma}_{xx}^{-1}(\hat{\Sigma}_{xx} + \hat{\Sigma}_{ff})\hat{B}_{\text{LSE}},$$

when $\hat{\Sigma}_{xx}$ is nonsingular.

Proof. The representation (2.5), the fact that the reparameterization (2.3) is one–to–one onto, and the given assumption that restrictions are placed only on Σ_{xx} and Σ_{ff} enable us to find maximum likelihood estimates of μ, Σ_{xx} and Σ_{ff} by maximizing the marginal likelihood of $X_1, \ldots, X_n$. Maximizing first over μ yields $\hat{\mu} = \overline{X}$. It then follows that $\hat{\Sigma}_{xx}$ and $\hat{\Sigma}_{ff}$ depend only on W_{XX} (as does $\hat{\Lambda}$). Similarly, maximum likelihood estimators of $\xi, \boldsymbol{\beta}$ and $\Sigma_{YY\cdot X}$ are obtained from the conditional likelihood of the Y_i's given the X_i's. Standard theory yields

$$\hat{\alpha} = \overline{Y} - \overline{X}\hat{B}_{\text{LSE}}, \quad \hat{\beta} = \hat{B}_{\text{LSE}} = W_{XX}^{-1}W_{XY},$$

$$\hat{\Sigma}_{YY\cdot X} = \frac{1}{n}\left(W_{YY} - W_{YX}W_{XX}^{-1}W_{XY}\right).$$

The formulas for $\hat{B}, \hat{\alpha}$ and $\hat{\Sigma}_{ee}$ follow directly from (2.4) and the definition of $\hat{\Lambda}$. It is conceivable that $\hat{\Sigma}_{xx}$ can be singular; in this case, maximum likelihood estimators of B, α and Σ_{ee} do not exist. $\square$

Examples of restrictions that may be placed on Σ_{xx} and Σ_{ff} to identify these parameters are given in Fuller (1987); see also Fuller and Hidiroglou (1978). It should be mentioned that Theorem 2.1 allows for more than $\frac{1}{2}r(r+1)$ functionally independent restrictions. For example, Theorem 2.1 applies when Σ_{xx} and Σ_{ff} are assumed to be known, or known up to a constant multiple. However, most investigators would prefer to impose a minimal number of restrictions, so that as little a'priori information as possible is required.

It is well known that $E(\hat{B}_{\mathrm{LSE}}|W_{XX}) = \boldsymbol{\beta}$. Theorem 2.1 states that $\hat{\Lambda}$ is a function of the data only through W_{XX}. Thus, if $E(\hat{B}_{\mathrm{MLE}})$ exists,

$$\begin{aligned} E\left[\hat{B}_{\mathrm{MLE}}\right] &= E\left\{E\left[\hat{\Lambda}^{-1}\hat{B}_{\mathrm{LSE}}|W_{XX}\right]\right\} \\ &= E\left[\hat{\Lambda}^{-1}\boldsymbol{\beta}\right] = E\left[\hat{\Lambda}^{-1}\Lambda\right]B. \end{aligned} \tag{2.7}$$

However, it is frequently the case that $E(\hat{B}_{\mathrm{MLE}})$ fails to exist. From the above analysis and Fubini's Theorem, the failure of $E(\hat{B}_{\mathrm{MLE}})$ to exist must stem from the failure of $E(\hat{\Lambda}^{-1})$ to exist. That is

$$E[\hat{B}_{\mathrm{MLE}}] \text{ exists if and only if } E[\hat{\Lambda}^{-1}] \text{ exists.} \tag{2.8}$$

When $E[\hat{B}_{\mathrm{MLE}}]$ fails to exist, the result (2.8) indicates that $\hat{\Lambda}^{-1}$ should be replaced in (2.6) with an alternative estimator which has smaller tails in its distribution. An example of an adjustment of $\hat{B}_{\mathrm{MLE}}$ to obtain better moment properties appears in Chapter 2 of Fuller (1987). However, this adjustment is not based on the perspective provided by the result (2.8).

If interest is centered on estimating the slope matrix B of the regression of the dependent variables on the true predictors x_i, Theorem 2.1 shows that B can be estimated in two stages:

(1) Estimate Λ from data on $X_1, \ldots, X_n$,

(2) Use $\hat{\Lambda}$ to modify the classical least squares estimator $\hat{B}_{\mathrm{LSE}}$ obtained from the regression of $Y_1, \ldots, Y_n$ on $X_1, \ldots, X_n$.

When $r = 1$, step (2) is known to psychometricians as the "correction for attenuation" of $\hat{B}_{\mathrm{LSE}}$ (see Fuller, 1987).

3. THE ROLE OF THE RELIABILITY MATRIX Λ

Note that

$$(x_i, X_i) \sim \text{ MVN}\left((\mu, \mu), \quad \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx} \\ \Sigma_{xx} & \Sigma_{XX} \end{pmatrix}\right),$$

so that

$$x_i | X_i \sim \text{ MVN}\left(\mu(I_r - \Lambda) + x_i\Lambda, \quad \Sigma_{xx}(I_r - \Lambda)\right). \tag{3.1}$$

Thus, Λ is the matrix of slopes for the regression of the true predictors x_i on the observed predictors X_i. From this result, it is apparent that Λ must be nonsingular in order for X_i to carry information about all elements of x_i. Further, the magnitudes of the elements of Λ indicate the precision with which X_i estimates (predicts) x_i, and thus indicates the precision with which the slope matrix B can be estimated.

This can be seen another way. It follows from (3.1) that

$$x_i = \mu(I_r - \Lambda) + X_i\Lambda + s_i, \tag{3.2}$$

where

$$s_i \sim \text{ MVN}(\mathbf{0}, \ \Sigma_{xx}(I_r - \Lambda)$$

is independent of X_i. (The residual s_i is also independent of e_i since it is a function only of x_i and f_i.) Substitute (3.2) into the expression for Y_i in (1.1). Thus,

$$Y_i = \alpha + \mu(I_r - \Lambda)B + X_i\Lambda B + (e_i + s_iB). \tag{3.3}$$

Note that $e_i + s_iB$ is independent of X_i. The equation (3.3) shows that B is the slope matrix of the regression of Y_i on $X_i\Lambda$. The least squares estimator from the regression of Y_i on $X_i\Lambda$, it is exists, is known to be the componentwise minimum variance unbiased estimator of B. For this estimator to exist, the sample cross product matrix $\Lambda' W_{XX}\Lambda$ of the $X_i\Lambda$ must be nonsingular. Assuming that Σ_{XX} is nonsingular, $\Lambda' W_{XX}\Lambda$ is nonsingular almost surely if and only if Λ is nonsingular.

If Λ is singular, there exists a non–zero vector t for which $t\Lambda' = 0$. In this case, tB is not identifiable, and thus cannot be estimated.

If Λ is nonsingular, the least squares estimator of the slope matrix of the regression of Y_i on $X_i\Lambda$ is

$$(\Lambda' W_{XX}\Lambda)^{-1}\Lambda' W_{XY} = \Lambda^{-1} W_{XX}^{-1} W_{XY} = \Lambda^{-1}\hat{B}_{\text{LSE}}.$$

The estimator $\Lambda^{-1}\hat{B}_{\text{LSE}}$ of B can only be used when Λ is known, but its accuracy is certainly a lower bound to the accuracies of estimators of B (such as those of Theorem 2.1) used when Λ must be estimated.

For an $r \times p$ matrix A with rows $a_1, \ldots, a_r$, let

$$\text{vec}(\text{A}) = (a_1, a_2, \ldots, a_r).$$

Note that vec(A) is an rp–dimensional row vector. Given W_{XX}, the conditional covariance matrix of $\text{vec}(\hat{B}_{\text{LSE}})$ is known to be $W_{XX}^{-1} \otimes \Sigma_{YY\cdot X}$.

Since

$$E(W_{XX}^{-1}) = (n - r - 2)^{-1}\Sigma_{XX}^{-1},$$

and $\hat{B}_{\text{LSE}}$ is a conditionally (given W_{XX}) unbiased estimator of ΛB, it follows that the unconditional covariance matrix of $\text{vec}(\Lambda^{-1}\hat{B}_{\text{LSE}})$ is

$$\text{cov}\left[\text{vec}\left(\Lambda^{-1}\hat{B}_{\text{LSE}}\right)\right] = \left(\frac{1}{n-r-2}\right)\left\{\Lambda^{-1}\Sigma_{XX}^{-1}(\Lambda')^{-1} \otimes \Sigma_{YY\cdot X}\right\}. \qquad (3.4)$$

A Canonical Analysis

Equation (3.4) shows that the accuracy with which B can be measured (by $\Lambda^{-1}\hat{B}_{\text{LSE}}$, when Λ is known and nonsingular) is inversely related to the magnitude of Λ for fixed Σ_{XX} and $\Sigma_{YY\cdot X}$. However, $\Sigma_{YY\cdot X}$ also depends upon Λ. To obtain a more precise insight into how the accuracy of estimation of B depends on Λ, we can make use of the following canonical analysis.

Since Σ_{xx} is assumed to be nonsingular, there exists a nonsingular matrix T such that

$$\Sigma_{xx} = TT', \quad \Sigma_{ff} = TD_aT', \qquad (3.5)$$

where

$$D_a = \text{diag}\,(a_1, a_2, \ldots, a_r).$$

Note that

$$T^{-1}\Sigma_{xx}(T')^{-1} = I_r, \quad T^{-1}\Sigma_{ff}(T')^{-1} = D_a \tag{3.6}$$

is a simultaneous diagonalization of Σ_{xx} and Σ_{ff}. Further,

$$\Lambda = (T')^{-1}(I_r + D_a)^{-1}T' = (T')^{-1}D_\lambda T', \tag{3.7}$$

where $\lambda_i = (1 + a_i)^{-1}, i = 1, \ldots, r$, and D_λ=diag$(\lambda_1, \ldots, \lambda_r)$. From (3.7), we see that the columns of $(T')^{-1}$ are eigenvectors of Λ, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_r$.

Let Γ be the matrix of orthonormal eigenvectors of Σ_{ee}; thus

$$\Sigma_{ee} = \Gamma D_\omega \Gamma', \quad D_\omega = \text{diag}(\omega_1, \ldots, \omega_p), \tag{3.8}$$

where $\omega_1, \ldots, \omega_p$ are the eigenvalues of Σ_{ee}.

Let

$$\begin{aligned} \tilde{Y}_i &= Y_i\Gamma = \tilde{\alpha} + \tilde{x}_i\Delta + \tilde{e}_i, \\ \tilde{X}_i &= X_iT^{-1} = \tilde{X}_i + \tilde{f}_i, \end{aligned} \tag{3.9}$$

where

$$\begin{aligned} \tilde{\alpha} &= \alpha\Gamma, \quad \tilde{x}_i = x_iT^{-1}, \quad \Delta = TB\Gamma, \\ \tilde{e}_i &= e_i\Gamma, \quad \tilde{f}_i = f_iT^{-1}. \end{aligned} \tag{3.10}$$

It follows from (1.1), (1.2), (1.3), (3.7), (3.8), (3.9) and (3.10) that

$$(\tilde{x}_i, \tilde{e}_i, \tilde{f}_i) \sim \text{MVN}\left((\tilde{\mu}, 0, 0), \begin{pmatrix} I_r & 0 & 0 \\ 0 & D_a & 0 \\ 0 & 0 & D_w \end{pmatrix}\right), \tag{3.11}$$

where $\tilde{\mu} = \mu T^{-1}$. Repeating our earlier analysis, we find that Δ can be estimated by $\hat{\Delta} = D_\lambda^{-1}\hat{\Delta}_{\text{LSE}}$ where

$$\hat{\Delta}_{\text{LSE}} = T\hat{B}_{\text{LSE}}\Gamma$$

is the least squares estimator of the slope matrix of the regression of $\tilde{Y}_i$ on $\tilde{X}_i, i = 1, \ldots, n$. Further, since

$$\Sigma_{\tilde{X}\tilde{X}} = I_r + D_a = D_{1+a} = D_\lambda^{-1},$$

$$\Sigma_{\tilde{Y}\tilde{Y}\cdot\tilde{X}} = D_\omega + \Delta'\Delta - \Delta'(I_r + D_a)^{-1}\Delta$$

$$= D_\omega + \Delta'\Delta - \Delta' D_\lambda \Delta,$$

it follows that

$$\operatorname{cov}\left[\operatorname{vec}(\hat{\Lambda})\right] = \left(\frac{1}{n-r-2}\right)\left\{D_\lambda^{-1} \otimes (D_\omega + \Delta'\Delta - \Delta' D_\lambda \Delta)\right\}. \tag{3.12}$$

Consider the (j, k)–th element Δ_{jk} of Δ. From (3.12) the variance of $\hat{\Delta}_{jk}$, the corresponding (j, k)–th element of $\hat{\Delta}$, is

$$\operatorname{var}(\hat{\Delta}_{jk}) = \left(\frac{1}{n-r-2}\right)(\lambda_j^{-1})\left(\omega_k + \sum_{i=1}^{r} \Delta_{ik}^2 - \sum_{i=1}^{r} \lambda_i \Delta_{ik}^2\right). \tag{3.13}$$

The reliability of the k–th element of $\tilde{Y}_i$ is, from (3.9) and (3.11), equal to

$$\tau_k = \frac{\operatorname{var}(\tilde{x}_i\Delta)_k}{\operatorname{var}(Y_i)_k} = \frac{\sum_{i=1}^{r} \Delta_{ik}^2}{\sum_{i=1}^{r} \Delta_{ik}^2 + \omega_k},$$

while λ_j is the reliability of the jth element of $\tilde{X}_i$. From (3.13),

$$\operatorname{var}(\hat{\Delta}_{jk}) = \left(\frac{1}{n-r-2}\right)\left(\sum_{i=1}^{r} \Delta_{ik}^2\right)\left(\frac{1}{\lambda_j \tau_k}\right)\left(1 - \tau_k \sum_{i=1}^{r} \lambda_i \left[\frac{\Delta_{ik}^2}{\sum_{\ell=1}^{r} \Delta_{\ell k}^2}\right]\right), \tag{3.14}$$

Thus, if the length of the jth row of Δ is fixed, and for a given reliability τ_k for the k–th component of the transformed vector $\tilde{Y} = Y\Gamma$ of observations on the dependent variables, the variance of $\hat{\Delta}_{jk}$ is a decreasing function of each of the eigenvalues λ_j of the reliability matrix Λ. Note from (3.10) that Δ_{jk} is a particular linear combination of the elements of B obtained from the jth row t_j of T and the k–th column γ_k' of Γ; that is, $\Delta_{jk} = t_j B \gamma_k'$.

Note. When $p = 1$, so that there is one dependent variable measured, the transformation $Y \to \tilde{Y} = Y\Gamma$ is trivial ($\Gamma = \pm 1$). The results above show that the variances of the estimators $\hat{\Delta}_j$ of the components Δ_j of the column vector $\Delta = TB$ are decreasing functions of the eigenvalues of Λ, for fixed Δ_j^2 and fixed reliability τ of the scalar measurement Y.

A canonical analysis of Λ (or of $\hat{\Lambda}$, when Λ is estimated) can clearly be useful in multivariate linear EIVR contexts. It is easily shown that

$$0 \leq \lambda_j \leq 1, \quad j = 1, 2, \ldots, r, \tag{3.15}$$

where $\lambda_j = 1$ indicates that the j–th component of TX_i measures the corresponding component of Tx_i without measurement error, and a value of λ_j close to zero indicates that the j–th component of Tx_i is poorly measured by the corresponding component of TX_i. If one or more of the eigenvalues λ_j are close to zero, it will not be possible to estimate certain linear combinations of the elements of the slope matrix B with accuracy. In such cases, investigators may want to consider deleting poorly measured predictor variables, or linear combinations of predictor variables (components of $\tilde{X} = XT^{-1}$), from the study. Alternatively, other estimators may be considered (such as the ordinary least squares estimator). One can also try to use instrumental variables for the poorly measured true predictors to improve the reliabilities.

It is important to note that when $r \neq 1$ the eigenvalues λ_j of Λ are not the reliabilities of the components of X_i, although they are reliabilities for certain linear combinations of these components. It is also true that unless $r = 1$, the j–th diagonal element Λ_{jj} of Λ is not the marginal reliability of the jth component of X_i. Marginal reliabilities refer to individual components measured in isolation. In contrast, Λ refers to *joint* measurement of these components.

4. CONCLUSION

The results in Sections 2 and 3 give considerable justification to the study of the reliability matrix Λ of error–prone measurements of predictor vectors in multivariate linear regression.

Since estimation of Λ can be done entirely from observations on the predictors, pilot studies involving these predictors, where possible, are recommended. Such studies, by use of the canonical approach of Section 3, can give warning of potential accuracy problems in estimating B. In particular, near–singularity of Λ can result in problems very similar to those of multicollinearity in classical linear regression. Because the "true design matrix" $\sum_{i=1}^{n}(x_i - \overline{x})'(x_i - \overline{x})$, where $\overline{x} = n^{-1}\sum_{i=1}^{n} x_i$, is not directly observable in multivariate linear EIVR problems, pilot studies (and estimates of Λ) are the best way to spot such problems. Failure to study measurement reliability in advance of experimentation can result in uselessly inaccurate data, and a waste of valuable experimental resources.

REFERENCES

Bickel, P. J. and Ritov, Y. (1987), "Efficient Estimation in the Errors in Variables Model," *Annals of Statistics*, **15**, 513–540.

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

Fuller, W. A. and Hidiroglou, M. A. (1978), *Journal of the American Statistical Association*, **73**, 99–104.

Gleser, L. J. (1983), "Functional, Structural and Ultrastructural Errors–in–Variables Models," *Proceedings Business and Economics Statistics Section*, American Statistical Association, 57–66.

Nussbaum, M. (1984), "An Asymptotic Minimax Risk Bound for Estimation of a Linear Functional Relationship," *Journal of Multivariate Analysis*, **14**, 300–314.