 Vol. 23: 257–274, 1996

The Precision of the Estimated Generalized Least Squares Estimator in Multivariate Calibration

ROLF SUNDBERG

Stockholm University

ABSTRACT. In multivariate calibration, based on multivariate linear regression of y on x, the so-called classical estimator of an unknown x from an observed y is the estimated generalized least squares (EGLS) estimator. The parameters involved in it are estimated from the calibration sample data. In this paper approximation formulae and approximately unbiased estimators for the variance–covariance matrices of the EGLS estimator and the simpler unweighted least squares estimator are given. Particularly noteworthy is the multiplicative contribution from the uncertainty in the estimated weighting matrix of EGLS. If dim y is not much smaller than the calibration sample size n this factor, being $(n - \dim x - 2)/(n - \dim y - 2)$ can be disastrous for the precision of the EGLS estimator. Further alternative estimators are discussed. Theory is supplemented by a simulation study in which the variance approximations show satisfactory accuracy even under bad conditions.

Key words: generalized least squares, least squares, variance estimation

1. Introduction

The classical multivariate calibration problem is concerned with two sets of variables, X and Y, where Y has a multivariate linear regression on X. This regression is estimated from a calibration or training sample and used to estimate ("predict") a fixed but unknown future X from a corresponding new observed Y. The future X and Y will here be denoted ξ and Z, respectively, to distinguish them from the data of the calibration sample. The q variables forming the response vectors Y and Z represent quick or cheap measurements on items for which X and ξ represent a p-dimensional characteristic that is laborious or expensive to measure. It is natural to insist that $q \geqslant p$, since otherwise ξ cannot be determined uniquely from Z even in the case of error-free measurements. All difficulties to be discussed arise when $q > p$, but this should not be taken as an argument for choosing $q = p$, since further components in Y (Z) might contribute additional information.

In applications X (ξ) may stand for the chemical composition of a laboratory sample, or more specifically the concentrations of p specified substances, and Y (Z) can be spectrophotometric instrument registrations of absorbance, emittance or reflectance of light at $q \geqslant p$ suitably selected wavelengths. Several examples from the IR and NIR spectral regions are found in the books by Martens & Næs (1989) and Brown (1993). The potential number of wavelengths to be used may be much larger than the calibration sample size, but often a moderately sized subset of representatives is selected, see for example Brown (1993, ch. 7).

Another type of example is when the measurement device to be calibrated is a wind tunnel balance, when X represents a $p = 6$-dimensional vector of forces and moments of forces, whereas Y represents a $q \geqslant 6$-dimensional vector of measured changes in electrical resistance of stretched filaments. Examples of biometric applications are as yet dominated by age determinations: estimation/prediction of the age of a young animal from q different body measures (Wood, 1982), of gestational age from fetal bone measurements by ultrasound (Oman & Wax, 1984), or of the age of a rhinoceros from its two horn lengths (Clarke, 1992). In such applications we typically have a small p and a moderate q-value.

The (estimated) *generalized least squares* estimator of ξ in the regression model is the classical estimator of ξ from an observed Z, proposed as early as the book by Williams (1959) and further discussed by Brown (1982, 1993) and others. The regression parameters and the weights matrix (inverse of residual variance matrix) are estimated from the calibration data, and the sample residual variance is assumed non-singular. The non-linear dependence on the estimated parameters makes it difficult to write down a formula for the variance of the estimator, or an unbiased estimator of this variance. The aim of the present paper is to provide approximate but practically useful such formulae. Also, the variance formula can be used to judge whether the estimated generalized least squares estimator will be better than the simple least squares estimator, that does not pretend to use the error variance matrix for optimal weighting. It turns out that the uncertainty in the weights matrix can have a disastrous effect on the variance of the generalized least squares estimator.

It might be asked why we should be interested in point estimates with approximate standard errors rather than confidence regions. Besides the argument that an approximate interval can be based on a combination of the approximate standard error with approximate normality, note that when $q > p$ there is no agreement on what confidence region otherwise to use. The exact one proposed by Brown (1982) has properties that speak strongly against it. A recent exact alternative by Mathew & Kasala (1994) is quite complicated, and the authors admit its shape is in general unknown and not easily analysed. Other regions, related to the use of the generalized least squares or simple least squares estimators, are either asymptotic, assuming calibration sample size $n \to \infty$ for fixed q, or conservative and relatively complicated to calculate (Wood, 1982; Fujikoshi & Nishii, 1984; Davis & Hayakawa, 1987; Brown & Sundberg, 1987; Oman, 1988; Mathew & Zha, 1996).

If we want a point estimate, we also need its theoretical variance, to guide us in the choice between different estimators, and its estimated standard error, to be given as a simple measure of precision together with the estimate itself. Hence, good approximations to the variance–covariance matrices of the estimated generalized least squares and simple least squares estimators, in the form of simple expressions that are easily evaluated and estimated, are helpful. Also, we want formulae valid even if n is not very much larger than q. We propose such formulae below, and we advocate their routine use for providing standard errors etc. for these two estimates.

2. The statistical model

The assumed multivariate linear regression model for the calibrating n-sample data is the standard one, written

$$Y = \mathbf{1}\alpha^{\mathrm{T}} + XB + E, \tag{2.1}$$

where Y now represents the whole $n \times q$ matrix of n row-wise arranged response vectors Y_i, X likewise a fixed $n \times p$ matrix of n explanatory vectors X_i, and E an $n \times q$ matrix of n row-wise arranged error vectors ε_i, $i = 1, \ldots, n$. The $\mathbf{1}$ stands for a column vector of ones, α and B are respectively a q-length column vector and a $p \times q$ matrix of parameters. The explanatory vectors X_i are treated as fixed, even though in reality they might have been more or less randomly generated. The error vectors ε_i are assumed mutually independent and normally distributed, $N_q(0, \Gamma)$. The normal distribution is not without importance for our results since we will use properties of the Wishart distribution.

The corresponding model for $Z(q \times 1)$ is

$$Z = \alpha + B^{\mathrm{T}}\xi + \varepsilon, \tag{2.2}$$

where ξ is regarded as a parameter, and where ε satisfies the same assumptions as ε_i and is normally distributed and independent of E. However, note that in applications it might not be obvious or even true that ε has the same covariance matrix Γ as the calibration sample errors ε_i.

We will mostly assume that there are m independent replicates of the measurement Z, with average $\bar{Z}$, that is

$$\bar{Z} = \alpha + B^{\mathrm{T}}\xi + \bar{\varepsilon}, \tag{2.3}$$

where the covariance matrix of $\bar{\varepsilon}$ is Γ/m. By doing so we can recognize the effect of the error ε in Z separated from the contribution of the calibration errors in subsequent formulae. In particular, to find out the consequences of a possibly different covariance matrix Γ_z for Z than the Γ for Y_i we just go into the formulae and change $(1/m)\Gamma$ to Γ_z.

Let $\hat{B}$ and $\hat{\alpha} = \bar{y} - \hat{B}^{\mathrm{T}}\bar{x}$ be the usual least squares estimators of B and α, that is

$$\bar{y} = Y^{\mathrm{T}}\mathbf{1}/n,$$

$$\bar{x} = X^{\mathrm{T}}\mathbf{1}/n,$$

and

$$\hat{B} = S_{xx}^{-1}(X - \mathbf{1}\bar{x}^{\mathrm{T}})^{\mathrm{T}}(Y - \mathbf{1}\bar{y}^{\mathrm{T}}), \tag{2.4}$$

with S_{xx} being the sum of centred squares and products matrix,

$$S_{xx} = (X - \mathbf{1}\bar{x}^{\mathrm{T}})^{\mathrm{T}}(X - \mathbf{1}\bar{x}^{\mathrm{T}}).$$

Let $\hat{\Gamma}(q \times q)$ be the residual sum of squares and products matrix of the calibration sample, divided by its appropriate number of degrees of freedom,

$$\hat{\Gamma} = (Y - \mathbf{1}\hat{\alpha} - X\hat{B})^{\mathrm{T}}(Y - \mathbf{1}\hat{\alpha} - X\hat{B})/(n - p - 1). \tag{2.5}$$

We make the standard assumption $n > p + q$, so that $\hat{\Gamma}$ is well-defined and non-singular with probability one. The statements of the following lemma are well-known, see for instance ch. 17 of Arnold (1981).

Lemma 1

The statistics $\bar{y}$, $\hat{B}$, $\hat{\Gamma}$ and $\bar{Z}$ are mutually independent, and

(1) *$\bar{y}$ is $N_q(\alpha + B^{\mathrm{T}}\bar{x}, \Gamma/n)$*

(2) *$\hat{B}$ is matrix normal, $N_{p,q}(B, S_{xx}^{-1}, \Gamma)$*

(3) *$\hat{\Gamma}$ is central Wishart, $W_q\{n - p - 1, \Gamma/(n - p - 1)\}$*

(4) *$\bar{Z} - \hat{\alpha} - \hat{B}^{\mathrm{T}}\xi = \bar{Z} - \bar{y} - \hat{B}^{\mathrm{T}}(\xi - \bar{x})$ is $N_q\left\{0, \left(\frac{1}{m} + \frac{1}{n} + (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})\right)\Gamma\right\}$.*

Property (4) follows from (1) and (2) by the following useful (and later used) calculation rule: If $\hat{B}$ is matrix normal as above and c is a constant p-vector, then the vector $\hat{B}^{\mathrm{T}}c$ is normal $N_q\{B^{\mathrm{T}}c, (c^{\mathrm{T}}S_{xx}^{-1}c)\Gamma\}$.

3. The least squares and estimated generalized least squares estimators

When $q = p$, so that $\hat{B}$ is a square, non-singular matrix, a natural estimator of ξ is obtained from the mean value relation in (4) of lemma 1 as

$$\hat{\xi} = (\hat{B}^{\mathrm{T}})^{-1}(\bar{Z} - \hat{\alpha}). \tag{3.1}$$

When $q > p$, (3.1) is easily generalized to the least squares (LS) estimator

$$\hat{\xi}_{\mathrm{LS}} = (\hat{B}\hat{B}^{\mathrm{T}})^{-1}\hat{B}(\bar{Z} - \hat{\alpha}). \tag{3.2}$$

Alternatively (3.2) may be written

$$\hat{\xi}_{\mathrm{LS}} = (\hat{B}^{\mathrm{T}})^{+}(\bar{Z} - \hat{\alpha}),$$

where B^{+} denotes the Moore–Penrose generalized inverse of B.

If Γ were known it could be used to form an estimator more efficient than (3.2), viz. the Γ^{-1}-weighted generalized least squares estimator

$$\hat{\xi}_{\mathrm{GLS}} = (\hat{B}\Gamma^{-1}\hat{B}^{\mathrm{T}})^{-1}\hat{B}\Gamma^{-1}(\bar{Z} - \hat{\alpha}). \tag{3.3}$$

Typically Γ is not known, but we may insert an estimate for it. When we choose the standard $\hat{\Gamma}$ of (2.5), based on the calibration residuals, we obtain what will here be called the estimated generalized least squares (EGLS) estimator,

$$\hat{\xi}_{\mathrm{EGLS}} = (\hat{B}\hat{\Gamma}^{-1}\hat{B}^{\mathrm{T}})^{-1}\hat{B}\hat{\Gamma}^{-1}(\bar{Z} - \hat{\alpha}). \tag{3.4}$$

The estimator (3.4) is the main object of study in this paper. Note that when $q = p$, (3.3) and (3.4) both simplify to (3.1). Even in the absence of replicates, that is $m = 1$, the EGLS estimator (3.4) is not the maximum likelihood (ML) estimator (unless $p = q$). Heuristically this is because there are $q - p$ pieces of information on Γ inherent in the residual vector $Z - \hat{\alpha} - \hat{B}^{\mathrm{T}}\hat{\xi}_{\mathrm{EGLS}}$. The ML estimator can hardly be recommended, however, see Brown & Sundberg (1987) or Brown (1993).

If $m > 1$ there is also information on Γ in the variation betwen Z replicates, that could easily be incorporated in a pooled $\hat{\Gamma}$. We will not do this here, mostly because we want to keep the situation simple. Also, we will not go into the complications when $\hat{\Gamma}$ is singular, but the reader is referred to Sundberg & Brown (1989) for a study of the EGLS estimator in this case.

It was shown by Nishii & Krishnaiah (1988) that the variance–covariance matrix $\mathrm{var}\,(\hat{\xi}_{\mathrm{EGLS}})$ exists if

(1) $q \geqslant p + 2$ and
(2) $n \geqslant \max(q + 3, p + q + 1)$

For the mean to exist it is sufficient that $q \geqslant p + 1$.

For the sequel, we could assume that conditions (1) and (2) be satisfied, but this would be unnecessarily restrictive. In fact the variances need not exist in order for the approximate variance formulae given below to hold, if only the latter are interpreted as the variances of the approximating normal distributions of the estimator $\hat{\xi}$ in question. Note that the estimators (3.2)–(3.4) are consistent as $n \to \infty$ or $\Gamma \to 0$, and the approximation will be good for small S_{xx}^{-1} (e.g. due to large n) or small Γ. In an Appendix we prove the following result.

Theorem 1

Approximate variance–covariance matrices for the LS, GLS and EGLS estimators are given by the formulae

$$\mathrm{var}\,(\hat{\xi}_{\mathrm{LS}}) \approx \left\{\frac{1}{m} + \frac{1}{n} + (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})\right\}(BB^{\mathrm{T}})^{-1}B\Gamma B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1} \tag{3.5}$$

$$\mathrm{var}\,(\hat{\xi}_{\mathrm{GLS}}) \approx \left\{\frac{1}{m} + \frac{1}{n} + (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})\right\}(B\Gamma^{-1}B^{\mathrm{T}})^{-1} \tag{3.6}$$

$$\mathrm{var}\,(\hat{\xi}_{\mathrm{EGLS}}) \approx \frac{n - p - 2}{n - q - 2}\left\{\frac{1}{m} + \frac{1}{n} + (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})\right\}(B\Gamma^{-1}B^{\mathrm{T}})^{-1}, \tag{3.7}$$

valid asymptotically as $\hat{B} \to B$ *in distribution, that is in particular as* $S_{xx}^{-1} \to 0$ *or* $\Gamma \to 0$. *More precisely, the relative errors of the* $(1/m) + (1/n)$ *and* $(\xi - \bar{x})^{\mathrm{T}} S_{xx}^{-1} (\xi - \bar{x})$ *parts of the formulae are respectively* $O(\delta^2)$ *and* $O(\delta)$, *where* $\delta^2 = \mathrm{tr}\,(\Gamma)\ \mathrm{tr}\,(S_{xx}^{-1})$ *represents the magnitude of* $\mathrm{var}\,(\hat{B})$.

Nishii & Krishnaiah (1988, th. 5) gave exact formulae for bias and second order moments of the EGLS estimator in the special case $p = 1$, $m = 1$, extending the corresponding results for the GLS estimator by Lieftinck-Koeijers (1988). These formulae are given in terms of moments for rational functions of Poisson variables. Their MSE result can be written in the following form, convenient for showing its relationship with (3.7):

$$\begin{aligned} \mathrm{MSE}(\hat{\xi}_{\mathrm{EGLS}}) = \frac{n-3}{n-q-2}\Bigg[\left(1+\frac{1}{n}\right)E\left\{\frac{S_{xx}}{2K+q-2}\right\} \\ + (\xi-\bar{x})^2 E\left\{\frac{2K}{(2K-q-2)(2K-q-4)}\right\}\Bigg] \\ + (\xi-\bar{x})^2 E\left\{\frac{(q-2)(q-4)}{(2K-q-2)(2K-q-4)}\right\}, \end{aligned} \tag{3.8}$$

where K is Poisson $\{\frac{1}{2} S_{xx} (B\Gamma^{-1}B^{\mathrm{T}})^{-1}\}$. Letting $E(K) \to \infty$ and making first order approximations of the expected values in terms of $1/E(K)$, we obtain the special version of (3.7) from (3.8). Even though we have the exact formula (3.8), formula (3.7) is useful as a simple and convenient approximation also when $p = 1$, good enough under normal operating conditions. But (3.7) has an even more important role: After substitution of estimates for the unknown quantities it gives an adequate (estimated) standard errors formula, see theorem 2 below.

Formula (3.7) for EGLS is not consistent with the approximation formula for the S_{xx}^{-1}-weighted mean squared error

$$E\{(\hat{\xi}_{\mathrm{EGLS}} - \xi)^{\mathrm{T}} S_{xx}^{-1} (\hat{\xi}_{\mathrm{EGLS}} - \xi)\}, \tag{3.9}$$

as given in Fujikoshi & Nishii (1986, th. 1, p. 272). This is because in their derivation they failed to observe one of the ξ-dependent contributions to the variance; for details see the Appendix, formula (A.10). As a byproduct of the simulations of section 4 below, the Fujikoshi & Nishii formula was clearly observed to fit less well than (3.7) when ξ was at a distance from $\bar{x}$.

Note that if the parameters α, B and Γ were all known, so that we had only the uncertainty in $\bar{Z}$ to bother about, the estimator (3.3) with the true values of α and B inserted would be exactly unbiased and have the variance

$$\frac{1}{m}(B\Gamma^{-1}B^{\mathrm{T}})^{-1}.$$

In each of (3.5)–(3.7) the three terms within braces { } are due to the uncertainty in $\bar{Z}$, $\bar{y}$ and $\hat{B}$ respectively, cf. property (4) of lemma 1. The scalar factor

$$\frac{n-p-2}{n-q-2} \tag{3.10}$$

comes from the uncertainty in the weights matrix, the residual-based $\hat{\Gamma}$. If dimensions p and q are small and n is large, the latter factor is not of much importance, but in practice n is often of moderate size and q may be large. If $n = 20$, $q = 10$, $p = 2$, say, which are reasonable values, this factor is 2, and far from negligible.

It can (of course) be shown that the difference between (3.5) and (3.6) is non-negative definite. Hence, if it were not for the factor (3.10) in (3.7), the variance formulae would have

shown EGLS to be superior to LS. The matrix difference cannot be characterized by only n, p and q, and it is not possible to say anything simple and general as to the relative efficiencies of the EGLS and LS estimators.

In order to exemplify the question of relative efficiency, let us here briefly look theoretically at heteroscedasticity effects, and for simplicity assume Γ diagonal and B^T proportional to a vector of ones. To compare the matrix factors (scalars when $p = 1$) is then simply to compare arithmetic and harmonic means. Let say half the number of variances in Γ have the value σ_1^2 and the others $\sigma_2^2 < \sigma_1^2$. For EGLS to be more efficient than LS when the variance inflation factor (3.10) has the value 4, say (as for example in the simulation study setup of sec. 4.1 below, where $p = 1$, $q = 10$, $n = 15$), we must have $\sigma_2^2/\sigma_1^2 \geqslant 13$. This means that the more noisy observations in this case would contribute only a trifle of information in comparison with the more precise ones.

When we use the LS or the EGLS estimator and want to estimate its precision it is tempting just to insert the estimated values for the unknown parameters in (3.5) and (3.7), respectively. For EGLS this would be systematically misleading, however. There is no problem in replacement of B by $\hat{B}$, nor in replacement of Γ by $\hat{\Gamma}$ in (3.5) for LS, since (3.5) is linear in Γ and $\hat{\Gamma}$ is an unbiased estimator of Γ, uncorrelated with $\hat{B}$. However, doing the same in (3.7) for EGLS might lead to a serious underestimation of the EGLS variance. For EGLS the uncertainty in $\hat{\Gamma}$ will require a bias correction factor of a size similar to (3.10), namely $(n-p-1)/(n-q-1) \geqslant 1$. The need for this bias-correcting factor was also noted by Fujikoshi & Nishii (1986), but surprisingly not in full consequence.

Theorem 2

An approximately unbiased estimator of the approximating variance (3.5) *for* $\hat{\xi}_{LS}$ *is obtained by simply inserting the parameter estimates in* (3.5), *whereas* (3.7) *for the variance of* $\hat{\xi}_{EGLS}$ *requires the estimator*

$$\widehat{\text{var}}\,(\hat{\xi}_{EGLS}) = \frac{n-p-1}{n-q-1}\,\frac{n-p-2}{n-q-2}\left\{\frac{1}{m}+\frac{1}{n}+(\hat{\xi}-\bar{x})^T S_{xx}^{-1}(\hat{\xi}-\bar{x})\right\}(\hat{B}\hat{\Gamma}^{-1}\hat{B}^T)^{-1}. \quad (3.11)$$

Proof. The dimension reduction of the $q \times q$ matrix $\hat{\Gamma}$ to the $p \times p$ matrix $(A\hat{\Gamma}^{-1}A^T)^{-1}$ through a full rank $p \times q$ coefficients matrix A preserves the Wishart property but reduces the degrees of freedom by $q-p$, here that is from $n-p-1$ to $n-q-1$, see e.g. th. 17.15 of Arnold (1981). This explains the bias-correcting factor, since the mean of a Wishart matrix is proportional to its number of degrees of freedom.

The $(1/m)+(1/n)$ parts of the estimators are in fact exactly unbiased estimators of $\text{var}\,(\hat{\xi}_{LS})$ and $\text{var}\,(\hat{\xi}_{EGLS})$ under the special circumstance that $\xi = 0$. This follows from inspection of the proof of theorem 1.

4. Simulation results

The approximate variance formulae and variance estimators for the LS and EGLS estimators have been tried in simulation studies with $p = 1$ and $p = 2$. Of primary interest is how well the variance estimators for EGLS and LS approximate the true variances or mean squared errors under various conditions. When $p = 1$ we can compare with the inconvenient but exact formulae for EGLS provided by Nishii & Krishnaiah (1988), and when additionally Γ happens to be proportional to I_q the exact MSE for LS is obtained from Lieftinck-Koeijers' (1988) formula for GLS.

4.1. Simulations with a 1-dimensional x, (p = 1)

The following setup is used.

(i) Fixed values $p = 1$, $q = 10$ and $n = 15$, $m = 1$. In particular, this implies that the variance factor (3.10), representing the contribution from the uncertainty in the EGLS weights matrix, has the high value 4. To form the EGLS variance estimator we multiply also by $(n - p - 1)/(n - q - 1) = 3.25$, according to (3.10).

(ii) An X design with $\bar{x} = 0$, $S_{xx} = 14$, and ξ-values $\xi = 0$ and $\xi = 2$ to be estimated. Hence, the empirical standard deviation of the calibration x-values is $(S_{xx}/(n-1))^{1/2} = 1$, so $\xi = 2$ represents quite an unfavourable case, that is chosen here to make the uncertainty in $\hat{B}$ have a substantial influence on var $(\hat{\xi})$. In practice, if we expected ξ-values of that size we should certainly ensure much more spread in the calibration sample. Anyhow, with this setup the scalar factor common to the LS and EGLS variances is

$$\frac{1}{m} + \frac{1}{n} + \frac{(\xi - \bar{x})^2}{S_{xx}} = 1 + \frac{1}{15} + \frac{2^2}{14} \approx 1.35,$$

so it is clear that a quarter of the variance derives from the uncertainty in the estimated regression lines.

(iii) A fixed regression coefficients B-vector of unit length,

$$B = 0.2\ (1\ 1\ 2\ 2\ 2\ 1\ 1\ 1\ 2\ 2).$$

However, note that the specific values and the ordering of the components is of some importance only when the noise is correlated and/or heteroscedastic.

(iv) A noise covariance matrix Γ that is either taken proportional to the identity, $\Gamma = \sigma^2 I_q$, or to have a correlated structure, $\Gamma = \sigma^2 C$, with the elements of C being corr $(y_j, y_k) = 2^{-|j-k|}$. As values of σ are taken $\sigma = 0.1$, $\sigma = 0.2$, and $\sigma = 0.4$. Even the smallest of these values is a large one, considering the unit standard deviation of the calibration x-values and the regression coefficient values 0.2 and 0.4. As will be seen from Table 1, in the most extreme case $\hat{\xi}_{\text{EGLS}}$ has a standard error s.e. $(\hat{\xi}_{\text{EGLS}}) > 1$, that is $\hat{\xi}_{\text{EGLS}}$ has an uncertainty greater than the spread in the calibration x-sample. These high σ-values are chosen in order to be able to illustrate that the true MSE and variance of $\hat{\xi}_{\text{EGLS}}$ do not increase exactly proportional to σ^2, whereas the approximation does so.

The results of 40 000 simulations by SAS:IML® are found in Tables 1 and 2 for each of the parameter combinations specified under (ii) and (iv) above. The variance formulae (3.7) and (3.5) for EGLS and LS respectively are used with the true parameters inserted. The formula value is compared with the simulation sample mean of the corresponding variance estimator, for EGLS given by (3.11). The sample variance of the variance estimate yields the standard error of this sample mean. It turns out that the relative standard error stays constant when parameters are changed, so this measure is given in the column head to indicate precision. The formula and its estimate are also compared with the actual simulation sample mean squared error of $\hat{\xi}$. A standard error for this statistic has been calculated by use of the 4th moment of the sample of $\hat{\xi}$-values (a considerable over-spread was seen, as compared with a χ^2 distribution).

The simulations use the same random number seed for all combinations of parameters tried. This means that all rows in Tables 1 and 2 below are based on the same random samples, and this facilitates comparisons between rows. Some parallel runs have been carried out with other seeds, confirming that the results shown are representative.

Table 1. *True and estimated precision of EGLS when p = 1*

Noise size and structure	ξ	True MSE (3.8)	Empirical MSE($\hat{\xi}_{EGLS}$), s.e. = 1.2	Formula (3.7) w. true par.	Av. est. (3.11) w. true ξ, s.e. = 0.4	Av. est. (3.11) w. $\hat{\xi}_{EGLS}$, s.e. = 0.4	Bias^2 of $\hat{\xi}_{EGLS}$ in % of MSE
Uncorrel.	$\xi = 0$	0.0425	+0.3%	+0.4%	+0.3%	+0.6%	0
w. $\sigma = 0.1$	$\xi = 2$	0.0539	−0.9%	+0.3%	+0.2%	+0.3%	0.2%
Correlated	$\xi = 0$	0.0881	−0.4%	+0.9%	+0.4%	+1.0%	0
w. $\sigma = 0.1$	$\xi = 2$	0.1119	−0.9%	+0.7%	+0.2%	+0.4%	0.5%
Uncorrel.	$\xi = 0$	0.168	+0.4%	+1.7%	+0.2%	+1.4%	0
w. $\sigma = 0.2$	$\xi = 2$	0.213	−0.1%	+1.4%	−0.1%	+0.2%	0.9%
Correlated	$\xi = 0$	0.343	−0.4%	+3.6%	+0.3%	+2.7%	0
w. $\sigma = 0.2$	$\xi = 2$	0.438	−1.0%	+2.8%	−0.4%	+0.4%	1.9%
Uncorrel.	$\xi = 0$	0.638	−0.3%	+7.0%	+0.2%	+4.6%	0
w. $\sigma = 0.4$	$\xi = 2$	0.820	−0.7%	+5.5%	−1.2%	+0.3%	3.5%
Correlated	$\xi = 0$	1.238	−1.3%	+14.9%	+0.2%	+9.0%	0
w. $\sigma = 0.4$	$\xi = 2$	1.614	−1.3%	+11.7%	−2.6%	+0.4%	6.8%

Column 1 specifies the noise structure: The combination of size (three different σ-values) and correlation structure ($\Gamma = \sigma^2 I_q$ or $\Gamma = \sigma^2 C$), as described under (iv).
Column 2 specifies the ξ-value: central ($\xi = 0$), or in the tails ($\xi = 2$).
Column 3 gives the true mean squared error MSE($\hat{\xi}_{EGLS}$), according to the exact formula (3.8) deriving from Nishii & Krishnaiah (1988). All subsequent columns are given as percentage deviations from these true MSE values.
Column 4 shows the empirical MSE($\hat{\xi}_{EGLS}$) from the 40 000 simulations. The relative standard error of the empirical MSE as an estimate of the true MSE is 1.2%, so for example, the first item in the column should be read $(+0.3 \pm 1.2)$%, representing value $\pm$s.e.
Column 5 gives the approximate variance according to formula (3.7), when the true parameter values are used for ξ, B and Γ.
Column 6 shows the average value of the estimate (3.11), when it is used with the true value for ξ instead of the estimate $\hat{\xi}_{EGLS}$. The relative standard error of this average is 0.4%.
Column 7 shows the average value of the estimate (3.11). The relative standard error of this average is 0.4% (only slightly higher than for column 6).
Column 8 shows the size of the squared bias according to the Nishii & Krishnaiah (1988) exact formula, but here given in % of the true MSE.

The exact values of the MSE of $\hat{\xi}_{EGLS}$, see (3.8), involving expected values of rational functions of Poisson Po (λ) variables with high λ values, have been computed by the Mathematica® software.

Conclusions from Table 1.

(1) From the "True MSE" column it is seen that doubling the value of σ, which makes formula (3.7) four times higher, yields a slightly more than four times increased MSE value when σ is as high as here. Changing Γ from $\sigma^2 I_q$ to $\sigma^2 C$ essentially doubles the MSE. This is because parameters were chosen such that $BC^{-1}B^T \approx 0.5BB^T$. Note that trends and tendencies in the columns of Table 1 clearly depend on the MSE value but not noticeably on whether Γ is correlated or not. Finally, for each choice of Γ a change of ξ from 0 to 2 should by theorem 1 yield an MSE value approximately 27% higher, since

$$\frac{1 + 1/15 + 2^2/14}{1 + 1/15} \approx 1.27.$$

This fits almost perfectly for the smaller MSE-values. When $\sigma = 0.4$ there is a hardly noticeable deviation from 27%; worst case 30%.

(2) Empirical MSE fits true MSE well.

(3) Formula (3.7), which will only be used at a planning stage, fits quite well for moderate imprecision, and when it fits less well for extreme imprecision it exaggerates, so it is conservative. Note that for fixed ξ the percentage deviation is essentially proportional to the MSE itself, or more precisely stated, proportional to $1/E(K)$ in accordance with a second order expansion of (3.8). At an MSE of magnitude $\leqslant 0.1$ the error is not more than 1%, which is certainly negligible for planning purposes.

(4) The variance estimate with true ξ will never be used in practice, but is included here for comparison. Significant deviations from zero occurs precisely when $\hat{\xi}_{\text{EGLS}}$ has an appreciable bias, more precisely the variance estimates with true ξ compensate to some extent but not fully for the squared bias in the MSE, by exaggerating the variance. If we had also included a column for the corresponding formula with the true Γ as well, it would have been seen that the values were essentially the same, but their relative standard error much smaller. This finding confirms what one can expect, viz. that the estimation of Γ also brings much inaccuracy into the precision estimate (3.11).

(5) The penultimate column is the most important one, since it shows the average performance of the estimator (3.11). When $\xi = 0$, (3.11) has a tendency to be conservative, like formula (3.7) but clearly to a lesser degree. When $\xi = 2$ this tendency disappears, evidently because we compare with an MSE that has a compensating squared bias component. A comparison with the previous column reveals explainable systematic differences: Since the estimate (3.11) is a quadratic function of ξ, it will have a higher average than the corresponding expression with the true ξ, and the higher the variation in ξ, the greater the difference between columns.

(6) How about the worst value of the last column, $\text{bias}^2 = 6.8\%$? In this case, when expressed in absolute values, for estimating $\xi = 2$ we have $\text{bias}(\hat{\xi}) = -0.35$ and

Table 2. *True and estimated precision of LS when $p = 1$*

Noise size and structure	ξ	True MSE	Empirical MSE($\hat{\xi}_{\text{LS}}$), s.e. = 0.7	Formula (3.5) w. true par.	Av{Var. est.} w. $\hat{\xi}_{\text{LS}}$ s.e. = 0.2	$(\text{Bias})^2$ of $\hat{\xi}_{\text{LS}}$ in % of MSE
Uncorrel.	$\xi = 0$	0.0106	−0.7%	+0.4%	+0.1%	0
w. $\sigma = 0.1$	$\xi = 2$	0.0136	−0.7%	−0.2%	−0.7%	1.0%
Correlated	$\xi = 0$	Not known	0.0257	+0.9%	+0.8%	0
w. $\sigma = 0.1$	$\xi = 2$	Not known	0.0326	+1.0%	+0.9%	≈0.2%
Uncorrel.	$\xi = 0$	0.0419	−0.7%	+1.7%	+0.3%	0
w. $\sigma = 0.2$	$\xi = 2$	0.0545	−0.6%	−0.8%	−2.9%	3.7%
Correlated	$\xi = 0$	Not known	0.102	+2.0%	+1.2%	0
w. $\sigma = 0.2$	$\xi = 2$	Not known	0.129	+2.4%	+1.4%	≈0.7%
Uncorrel.	$\xi = 0$	0.159	−0.7%	+7.0%	+1.1%	0
w. $\sigma = 0.4$	$\xi = 2$	0.222	−0.6%	−2.6%	−10.5%	13.1%
Correlated	$\xi = 0$	Not known	0.388	+7.0%	+3.1%	0
w. $\sigma = 0.4$	$\xi = 2$	Not known	0.486	+8.4%	+3.6%	≈3.1%

Columns are the same as for EGLS in Table 1, except that its column 6 has been omitted.
Column 3 contains the true MSE when a formula is available for its computation (i.e. when $\Gamma = \sigma^2 I_q$), and then all subsequent columns give their values as relative deviations from this value. Else (i.e. when $\Gamma = \sigma^2 C$) the values are given as relative deviations from the noise-affected empirical MSE of column 4.

var $(\hat{\xi}) = 1.5$, so we should not be happy with the calibration situation as such. In this perspective a few per cent of systematic error in an estimate of precision is definitely not the problem.

The overall impression and the main message from Table 1 is that all variance and MSE values agree very well with the true MSE, except when the MSE is extremely large. In these extreme cases agreement is not excellent but relatively satisfactory.

Conclusions from Table 2.
As expected from theorem 1, the true MSE values for LS when $\Gamma = \sigma^2 I_q$ are only a quarter of the corresponding values for EGLS. The bias for $\xi = 2$ in this case might appear much larger for LS than for EGLS, but this is only when measured relative to the MSE; in absolute terms the two biases are identical. (*Remark*: apparently the bias is much lower when $\Gamma = \sigma^2 C$, but a closer study of this phenomenon is outside the scope of the present paper.) An influence of the higher squared bias percentage is clearly seen in the penultimate, average estimated variance column. If we remember that all values relate to the MSE we may conclude that the LS variance is at worst only slightly overestimated; nothing to worry about. If the LS estimator is very imprecise, however, its variance is not all, but its bias must also be judged.

The overall impression and the main message for LS from Table 2 is similar to that for EGLS from Table 1: The variance and average variance estimates for LS agree well with the true MSE when the MSE itself is small or moderately large.

4.2. *Simulations with a 2-dimensional x, (p = 2)*

The following setup is used:

(i) Fixed values $p = 2$, $q = 10$ and $n = 17$, $m = 1$. The variance factor (3.10) now has the value 2.6.

(ii) An orthogonal X design with $\bar{x} = 0$, $S_{xx} = 16 I_2$, and the three ξ-values $\xi = \{0, 0\}$, $\xi = \{2, 2\}$ and $\xi = \{2, -2\}$ to be estimated. The two latter points represent unfavourable values in both components simultaneously. The common scalar factor of the LS and EGLS variances is

$$\frac{1}{m} + \frac{1}{n} + \xi^{\mathrm{T}} S_{xx}^{-1} \xi = 1 + \frac{1}{17} + \frac{8}{16} \approx 1.56,$$

so for these two ξ more than a third of the variance derives from the uncertainty in the estimated regression lines. Note that $\xi = \{2, 2\}$ and $\xi = \{2, -2\}$ need not have the same precision estimates, even though their theoretical values from theorem 1 are the same.

(iii) Regression coefficients matrix with rows of length 1, given by

$$B = 0.2 \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 2 & 2 \\ 2 & 2 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 \end{pmatrix}, \quad BB^{\mathrm{T}} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

In the terminology of analytical chemistry this B-matrix shows a fairly bad selectivity between the two components: For none of the responses do the regression coefficients differ by more than a factor of 2.

(iv) The same noise covariance matrices $\Gamma = \sigma^2 I_q$ and $\Gamma = \sigma^2 C$ as for $p = 1$.

As for $p = 1$, 40 000 simulations have been carried out using SAS:IML®, and the same statistics are calculated from the simulations, except that the variance is no longer a scalar,

but a 2×2 matrix. Its two diagonal elements will be averaged over, since they are exchangeable by design. As regards the correlations, we should expect the following values from formulae (3.5) and (3.7):

When $\Gamma \propto I_q$: correlation -0.8 for both EGLS and LS.

When $\Gamma \propto C$: correlation ≈ -0.50 for EGLS and ≈ -0.53 for LS.

Both the actual simulated correlations and the average estimated correlations agree quite well with these figures. We only make this remark here and do not give any correlation values below.

Finally note that when $p = 2$ there are no exact true values to compare with.

Conclusions from Table 3.

The actual MSE values are somewhat higher here than in Table 1, even though factor (3.10) is smaller. This should not come as a surprise, since we are now estimating a two-dimensional ξ.

There is a clear difference between the points $\{2, 2\}$ and $\{2, -2\}$. Nor should this be a surprise, considering the correlation between the two components of $\hat{\xi}$.

Generally, the simulation results indicate that formula (3.7) would be quite adequate for planning purposes, and the systematic errors in the variance estimate are much too small to be of any importance. Furthermore they are of the correct sign to give the variance estimates a tendency to be conservative. At the high MSE-values appearing when $\sigma \geqslant 0.2$, an eye should be kept on the possibility of a serious bias, however.

Table 3. *Simulated and estimated precision of EGLS when $p = 2$*

Noise size and structure	ξ	Empirical MSE($\hat{\xi}_{\text{EGLS}}$), rel. s.e. $< 0.9\%$	Formula (3.7) w. true par.	Av. est. (3.11) w. true ξ, (s.e. < 0.3)	Av. est. (3.11) w. $\hat{\xi}_{\text{EGLS}}$, (s.e. < 0.3)	(Bias)^2 of $\hat{\xi}_{\text{EGLS}}$ in % of MSE
Uncorrel.	$\xi = \{0, 0\}$	0.075	+1.8%	+0.7%	+1.6%	0
w. $\sigma = 0.1$	$\xi = \{2, 2\}$	0.110	+1.9%	+0.9%	+1.6%	$<0.1\%$
	$\xi = \{2, -2\}$	0.112	+0.8%	−0.2%	−0.4%	$\approx 1.8\%$
Correlated	$\xi = \{0, 0\}$	0.075	+1.5%	+0.9%	+1.8%	0
w. $\sigma = 0.1$	$\xi = \{2, 2\}$	0.111	+1.7%	+1.1%	+1.7%	$\approx 0.1\%$
	$\xi = \{2, -2\}$	0.112	+0.8%	+0.2%	+0.2%	$\approx 1.3\%$
Uncorrel.	$\xi = \{0, 0\}$	0.29	+5.3%	+0.2%	+3.7%	0
w. $\sigma = 0.2$	$\xi = \{2, 2\}$	0.43	+5.5%	+0.3%	+3.3%	$\approx 0.1\%$
	$\xi = \{2, -2\}$	0.44	+1.9%	−3.1%	−3.5%	$\approx 6.5\%$
Correlated	$\xi = \{0, 0\}$	0.29	+3.9%	+0.5%	+4.1%	0
w. $\sigma = 0.2$	$\xi = \{2, 2\}$	0.43	+4.3%	+0.8%	+3.1%	$\approx 0.5\%$
	$\xi = \{2, -2\}$	0.44	+2.0%	−1.3%	−1.3%	$\approx 4.5\%$

Columns are the same as for $p = 1$ in Table 1, except for the missing "true MSE" column. In each row are given the averages of the two diagonal variance elements. All percentages refer to deviations from the empirical MSE values, but note that these are affected by randomness (s.e. 0.9% in each ξ-component). It is not by misprints but by coincidence that $\Gamma = \sigma^2 I_q$ and $\Gamma = \sigma^2 C$ give almost identical MSE values.

Conclusions from Table 4.

Table 4 is seen to conform well with a combination of the findings in Tables 2 and 3. For example, as in the comparison of Table 2 with Table 1, the bias percentages for LS in Table 4 are higher than for EGLS in Table 3, but measured on an absolute scale biases are similar for LS and for EGLS. Also, the few high values in the penultimate column of averaged

Table 4. *Simulated and estimated precision of LS when $p = 2$*

Noise size and structure	ξ	Empirical MSE($\hat{\xi}_{LS}$), rel. s.e. $< 0.7\%$	Formula (3.7) w. true par.	Av{Var. est.} w. $\hat{\xi}_{LS}$, (s.e. < 0.2)	(Bias)^2 of $\hat{\xi}_{LS}$ in % of MSE
Uncorrel.	$\xi = \{0, 0\}$	0.030	-0.9%	-1.9%	0
w. $\sigma = 0.1$	$\xi = \{2, 2\}$	0.044	-1.1%	-2.2%	$<0.1\%$
	$\xi = \{2, -2\}$	0.044	-2.6%	-4.6%	$\approx 4.2\%$
Correlated	$\xi = \{0, 0\}$	0.033	-1.5%	-1.4%	0
w. $\sigma = 0.1$	$\xi = \{2, 2\}$	0.048	-1.6%	-1.6%	$<0.1\%$
	$\xi = \{2, -2\}$	0.048	-2.0%	-2.8%	$\approx 2.4\%$
Uncorrel.	$\xi = \{0, 0\}$	0.11	$+3.2\%$	-0.8%	0
w. $\sigma = 0.2$	$\xi = \{2, 2\}$	0.17	$+3.2\%$	-1.4%	$\approx 0.2\%$
	$\xi = \{2, -2\}$	0.18	-5.5%	-12.8%	$\approx 14.9\%$
Correlated	$\xi = \{0, 0\}$	0.13	-0.2%	-0.2%	0
w. $\sigma = 0.2$	$\xi = \{2, 2\}$	0.19	-0.2%	-0.4%	$\approx 0.1\%$
	$\xi = \{2, -2\}$	0.20	-4.0%	-7.1%	$\approx 8.8\%$

Columns are the same as for $p = 1$ in Table 2, except for the missing "true MSE" column. In each row are given the averages of the two diagonal variance elements. All percentages refer to deviations from the empirical MSE values, but note that these are affected by randomness (s.e. 0.7% in each of the two ξ-components).

variance estimates correspond to the high squared bias values, that is those cases when the MSE is clearly larger than the variance. As a general conclusion from Table 4, even when there is so much noise in data as in these simulations, the variance of $\hat{\xi}_{LS}$ is on the average well estimated by (3.5) with estimated parameters. However the bias of $\hat{\xi}_{LS}$ may be disturbing if the noise level is high and ξ is far out in an unfortunate direction.

4.3. General conclusions from the simulations

The simulations, which assumed a relatively high noise level in data, indicate that the theoretical variance formulae and estimated standard errors of theorems 1 and 2 are fully adequate under normal operating conditions. In extreme situations with high variance, however, the bias can be so large that the variance values alone are misleading, but then the whole calibration situation is likely to be unsatisfactory.

These conclusions are not exclusively based on the simulation results given in Tables 1–4, but also on various additional simulations performed under similar conditions.

5. Alternative estimators

Above we have demonstrated and illustrated how much the uncertainty in $\hat{\Gamma}$ contributes to the variance of the EGLS estimator. In particular we saw that it may be worse to use $\hat{\xi}_{EGLS}$ with its estimated weighting matrix $\hat{\Gamma}^{-1}$ than to abstain from any particular weighting, as in $\hat{\xi}_{LS}$ (Above all, this was under the not necessarily satisfied assumption that Γ is the same in the calibration as in the subsequent use of the instrument.) However, $\hat{\xi}_{LS}$ is not the only alternative to $\hat{\xi}_{EGLS}$, and we will briefly discuss some other possibilities here. Each of these alternatives will work even if $n \leqslant p + q$, that is even in case $\hat{\Gamma}$ of (2.5) is singular.

A simple possibility is to use diag ($\hat{\Gamma}$) instead of $\hat{\Gamma}$, that is to behave as if the components of the error vector are mutually uncorrelated. The resulting estimator $\hat{\xi}$ will not be very

sensitive to the estimation errors in the diagonal elements. Some simulations indicate that this estimator is much more precise in the case of strong heteroscedasticity. However, it remains to find a useful formula and estimator for its variance.

Næs (1986) proposed a linear factor structure on the noise E. In chemical applications such a latent factor structure model element is often natural, since it can represent variation in concentration of one or several polluting substances that have not been calibrated for. In terms of Γ this means that with r latent factors Næs fitted the structure

$$\Gamma = UU^{\mathrm{T}} + D,$$

where U is a $q \times r$ coefficients matrix (making UU^{T} of rank r) and D is a diagonal "pure error" variance matrix. For the choice of dimension r, Næs tried both cross-validation on the calibration sample and minimum mean squared prediction error calculated on a special prediction set of (ξ, Z)-values. For $r = 0$ we have the previous method again, and for $r = q$ we recover the EGLS method.

Brown (1992, personal communication) tried a principal component analysis of $\hat{\Gamma}$, retaining only the largest eigenvalues. More specifically let $\hat{\Gamma} = P\Lambda P^{\mathrm{T}}$, with P orthogonal and Λ a diagonal matrix of eigenvalues λ of $\hat{\Gamma}$. In $\hat{\xi}_{\mathrm{EGLS}}$ we use $\hat{\Gamma}^{-1} = P\Lambda^{-1}P^{\mathrm{T}}$. The procedure now is to retain only the highest λ-values (smallest $1/\lambda$-values) and to eliminate the influence of the other principal components by inserting zeroes in Λ^{-1}. In Brown's example (the sugar example of his book; see Brown (1993, sec. 7.8) he chose the number of eigenvalues to be retained according to minimum mean squared prediction error calculated on a special prediction set of (ξ, Z)-values, but in other situations a cross-validation on the calibration sample might of course be employed.

Another possibility to reduce the risk for near-singularities in $\hat{\Gamma}$, and at the same time to reduce the variance (3.7), is to add a small diagonal matrix δI_q to $\hat{\Gamma}$. In this way a sort of ridge-type estimator is formed.

A different type of alternative is to select a suitable subset of the response variables. Brown (1982) suggested a "test of additional information" for such selection, but this question has not attracted so much interest as the corresponding one of selection of explanatory variables for multiple regression in the prediction approach of natural calibration, see Brown (1993).

Finally we must mention the possibility of using some statistic constructed as a predictor in a natural calibration model, where x and ξ are regarded as randomly generated. In such a model x is regressed on y by some regression methodology, to form a predictor. Theoretical comparisons of efficiency between estimators and predictors are sparse. Sundberg (1985) compares the efficiencies of the estimated best linear predictor (EBLP) and the generalized least squares estimator assuming a precise calibration. Oman & Srivastava (1996) and Srivastava (1995) extend this comparison to $\hat{\xi}_{\mathrm{GLS}}$ and $\hat{\xi}_{\mathrm{EGLS}}$ for $p = 1$, by deriving exact bias and MSE formulae analogous to those for GLS and EGLS by Lieftinck-Koeijers (1988) and Nishii & Krishnaiah (1988). For various regression methods suitable for handling colinearity and other problems with many explanatory variables (high q-value) better than EBLP, the reader is referred to the books by Brown (1993) and Martens & Næs (1989).

Acknowledgements

The author is grateful to a referee for helpful comments on a previous version of the simulation study, and acknowledges financial support by the Swedish Natural Science Research Council.

References

Arnold, S. F. (1981). *The theory of linear models and multivariate analysis.* Wiley, New York.

Brown, P. J. (1982). Multivariate calibration (with discussion). *J. Roy. Statist. Soc. Ser. B* **44,** 287–321.

Brown, P. J. (1993). *Measurement, regression, and calibration.* Oxford University Press, Oxford.

Brown, P. J. & Sundberg, R. (1987). Confidence and conflict in multivariate calibration. *J. Roy. Statist. Soc. Ser. B* **49,** 46–57.

Clarke, G. P. Y. (1992). Inverse estimates from a multiresponse model. *Biometrics* **48,** 1081–1094.

Davis, A. W. & Hayakawa, T. (1987). Some distribution theory relating to confidence regions in multivariate calibration. *Ann. Inst. Statist. Math.* **39,** 141–152.

Fujikoshi, Y. & Nishii, R. (1984). On the distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Math. J.* **14,** 215–225.

Fujikoshi, Y. & Nishii, R. (1986). Selection of variables in a multivariate inverse regression problem. *Hiroshima Math. J.* **16,** 269–277.

Gleser, L. J. & Olkin, I. (1982). Estimation for a regression model with an unknown covariance matrix. In *Proceedings, Sixth Berkeley Symposium,* **1,** 541–568.

Lieftinck-Koeijers, C. A. J. (1988). Multivariate calibration: a generalization of the classical estimator. *J. Multivariate Anal.* **25,** 31–44.

Martens, H. & Næs, T. (1989). *Multivariate calibration.* Wiley, Chichester.

Mathew, T. & Kasala, S. (1994). An exact confidence region in multivariate calibration. *Ann. Statist.* **22,** 94–105.

Mathew, T. & Zha, W. (1996). Conservative confidence regions in multivariate calibration. To appear in *Ann. Statist.* **24.**

Næs, T. (1986). Multivariate calibration using covariance adjustment. *Biometrical J.* **28,** 99–107.

Nishii, R. & Krishnaiah, P. R. (1988). On the moments of the classical estimates of the explanatory variables under a multivariate calibration model. *Sankhyā A* **50,** 137–148.

Oman, S. D. (1988). Confidence regions in multivariate calibration. *Ann. Statist.* **16,** 174–187.

Oman, S. D. & Srivastava, M. S. (1996). Exact mean squared error comparisons of the inverse and classical estimators in multi-univariate linear calibration. To appear in *Scand. J. Statist.*

Oman, S. D. & Wax, Y. (1984). Estimating fetal age by ultrasound measurements: an example of multivariate calibration. *Biometrics* **40,** 947–960.

Srivastava, M. S. (1995). Comparison of the inverse and classical estimators in multi-univariate linear calibration. *Commun. Statist. Ser. A* **24,** 2753–2767.

Sundberg, R. (1985). When is the inverse regression estimator MSE-superior to the standard regression estimator in multivariate controlled calibration situations? *Statist. Probab. Lett.* **3,** 75–79.

Sundberg, R. & Brown, P. J. (1989). Multivariate calibration with more variables than observations. *Technometrics* **31,** 365–371.

Williams, E. J. (1959). *Regression analysis.* Wiley, New York.

Wood, J. T. (1982). Estimating the age of an animal: an application of multivariate calibration. In *Proceedings of 11th International Biometrics Conference.*

Received June 1994, in final form November 1995

Rolf Sundberg, Mathematical Statistics, Stockholm University, S-106 91 Stockholm, Sweden.

Appendix: Proof of Theorem 1

We desire formulae for the variance–covariance matrices of estimators of type

$$\hat{\xi} = (\hat{B}W\hat{B}^{\mathrm{T}})^{-1}\hat{B}W(\bar{Z} - \hat{\alpha}), \tag{A.1}$$

where the (symmetric) weights matrix W stands for the random $\hat{\Gamma}^{-1}$ in the estimated $\hat{\xi}_{\mathrm{EGLS}}$, for the true Γ^{-1} in the ideal GLS estimator $\hat{\xi}_{\mathrm{GLS}}$, and for the identity matrix in $\hat{\xi}_{LS}$. First, note that we may replace $\hat{\alpha} = \bar{y} - \hat{B}^{\mathrm{T}}\bar{x}$ in (A.1) by $\bar{y}$ without changing the variance, since $\hat{B}^{\mathrm{T}}\bar{x}$ only contributes the constant $\bar{x}$ to the right hand side. As the next step in the calculation of this variance we condition on $\hat{B}$ and write

$$\mathrm{var}\,(\hat{\xi}) = E_{\hat{B}}\{\mathrm{var}\,(\hat{\xi} \mid \hat{B})\} + \mathrm{var}_{\hat{B}}\{E(\hat{\xi} \mid \hat{B})\} = \text{Term 1} + \text{Term 2}. \tag{A.2}$$

Case I. Non-random W.

Term 1

When W is fixed we immediately get

$$\text{var}(\hat{\xi} \mid \hat{B}) = \left(\frac{1}{m} + \frac{1}{n}\right)(\hat{B}W\hat{B}^{\mathrm{T}})^{-1}\hat{B}W\Gamma W\hat{B}^{\mathrm{T}}(\hat{B}W\hat{B}^{\mathrm{T}})^{-1}. \tag{A.3}$$

This exact formula evidently provides an (exactly) unbiased estimator for Term 1, even after insertion of the unbiased $\hat{\Gamma}$ for Γ, since $\hat{B}$ and $\hat{\Gamma}$ are independent (lemma 1). The approximation of its expected value obtained by simply inserting B for $\hat{B}$ in (A.3) has the relative error $O(\delta^2)$, where $\delta^2 = \text{tr}(\Gamma)\,\text{tr}(S_{xx}^{-1})$. In the special cases $W = I$ and $W = \Gamma^{-1}$ we obtain

$$\text{Term 1} \approx \begin{cases} \left(\frac{1}{m} + \frac{1}{n}\right)(BB^{\mathrm{T}})^{-1}B\Gamma B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1} & \text{for } W = I, \\ \left(\frac{1}{m} + \frac{1}{n}\right)(B\Gamma^{-1}B^{\mathrm{T}})^{-1} & \text{for } W = \Gamma^{-1} \end{cases}$$

Term 2

For Term 2 we first note that

$$E(\hat{\xi} \mid \hat{B}) = \bar{x} + (\hat{B}W\hat{B}^{\mathrm{T}})^{-1}\hat{B}WB^{\mathrm{T}}(\xi - \bar{x}) = \xi - (\hat{B}W\hat{B}^{\mathrm{T}})^{-1}\hat{B}W(\hat{B} - B)^{\mathrm{T}}(\xi - \bar{x}). \tag{A.4}$$

This is a non-linear function of $\hat{B}$ and to approximate its variance we use propagation of errors technique, utilizing the fact that $\hat{B}$ is close to B. Since one factor, $\hat{B} - B$, has expected value zero, the contributions to the variance from all other $\hat{B}$ in (A.4) are relatively negligible. From lemma 1 we know the distribution of $\hat{B}$, and it follows that the vector

$$(\hat{B} - B)^{\mathrm{T}}(\xi - \bar{x})$$

is distributed as

$$N_q\{0, (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})\Gamma\}.$$

Using this result on (A.4) we obtain the variance approximation, of relative error $O(\delta)$,

$$\text{Term 2} = \text{var}_{\hat{B}}\{E(\hat{\xi} \mid \hat{B})\} \approx \begin{cases} (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})(BB^{\mathrm{T}})^{-1}B\Gamma B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1} & \text{for } W = I, \\ (\xi - \bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi - \bar{x})(B\Gamma^{-1}B^{\mathrm{T}})^{-1} & \text{for } W = \Gamma^{-1}. \end{cases}$$

Case II. $W = \hat{\Gamma}^{-1}$.

We proceed by first calculating the expectation and variance of $\hat{\xi}_{\text{EGLS}}$ conditional on $\hat{B}$, and to achieve this we will use properties of partitioned Wishart matrices. As a preparation we follow Gleser & Olkin (1972) and Fujikoshi & Nishii (1986) and start by introducing a version of the singular value decomposition of $\hat{B}$, to write

$$\hat{B} = ML_1Q^{\mathrm{T}} = MQ_1^{\mathrm{T}},$$

where M is a non-singular $p \times p$ matrix,

$$L_1 = (I_p \quad O),$$

with I_p being the $p \times p$ identity matrix and O a $p \times (q - p)$ zero matrix, and finally

$$Q = (Q_1 \quad Q_2)$$

is an orthogonal $q \times q$ matrix with $q \times p$ and $q \times (q-p)$ submatrices Q_1 and Q_2 of orthogonal columns. We introduce $\hat{\Gamma}_Q = Q^T \hat{\Gamma} Q$ and note that for given $\hat{B}$, $\hat{\Gamma}_Q$ is Wishart distributed, $W_q\{n-p-1, \Gamma_Q/(n-p-1)\}$. Furthermore

$$\hat{\Gamma}_Q^{-1} = Q^T \hat{\Gamma}^{-1} Q,$$

since Q is orthogonal, and hence

$$(\hat{\Gamma}_Q^{-1})_{11} = Q_1^T \hat{\Gamma}_Q^{-1} Q_1, \quad (\hat{\Gamma}_Q^{-1})_{12} = Q_1^T \hat{\Gamma}_Q^{-1} Q_2,$$

where $(\cdot)_{11}$ denotes the $p \times p$ upper left submatrix, and analogously for $(\cdot)_{12}$. With these notations we can write

$$\begin{aligned}
\hat{\xi}_{\text{EGLS}} - \bar{x} &= (MQ_1^T \hat{\Gamma}^{-1} Q_1 M^T)^{-1} MQ_1^T \hat{\Gamma}^{-1} QQ^T(\bar{Z} - \bar{y}) \\
&= (M^T)^{-1}(\hat{\Gamma}_Q^{-1})_{11}^{-1}((\hat{\Gamma}_Q^{-1})_{11} \quad (\hat{\Gamma}_Q^{-1})_{12}) Q^T(\bar{Z} - \bar{y}) \\
&= (M^T)^{-1}(I_p \quad (\hat{\Gamma}_Q^{-1})_{11}^{-1}(\hat{\Gamma}_Q^{-1})_{12}) Q^T(\bar{Z} - \bar{y}) \\
&= (M^T)^{-1}(I_p \quad -(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1}) Q^T(\bar{Z} - \bar{y}) \qquad \text{(A.5)}
\end{aligned}$$

To obtain the last equality in (A.5) the matrix inversion formulae (see e.g. Arnold, 1981, app. A2)

$$(A^{-1})_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$$

$$(A^{-1})_{12} = -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1}$$

are applied with $A = \hat{\Gamma}_Q$.

Remember that for given $\hat{B}$, matrices M and Q are constant. All dependence of $\hat{\Gamma}$ in (A.5) is in the component

$$(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1} Q_2^T(\bar{Z} - \bar{y}) \qquad \text{(A.6)}$$

Since $\hat{\Gamma}_Q$ is (conditionally) Wishart, $W_q(n-p-1, \Gamma_Q/(n-p-1)\}$, we know the joint distribution of $(\hat{\Gamma}_Q)_{22}$ and $(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1}$ from a basic result on partitioned Wishart matrices, in Arnold (1981) found as lem. 17.9. Here this result reads:

Lemma 2

(1) $(\hat{\Gamma}_Q)_{22}$ *is Wishart,* $W_{q-p}\{n-p-1, (\Gamma_Q)_{22}/(n-p-1)\}$;

(2) *Given* $(\hat{\Gamma}_Q)_{22}$, $(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1}$ *is conditionally matrix normal,*

$$N\{(\Gamma_Q)_{12}(\Gamma_Q)_{22}^{-1}, (\Gamma_Q^{-1})_{11}^{-1}, (\hat{\Gamma}_Q)_{22}^{-1}\}.$$

Given $\hat{B}$, $(\bar{Z} - \bar{y})$ and $(\hat{\Gamma}_Q)_{22}$, so that in particular $Q_2^T(\bar{Z} - \bar{y})$ is a constant column vector, lemma 2 implies that $(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1} Q_2^T(\bar{Z} - \bar{y})$ is (vector) normal,

$$N_p\{(\Gamma_Q)_{12}(\Gamma_Q)_{22}^{-1} Q_2^T(\bar{Z} - \bar{y}), (\bar{Z} - \bar{y})^T Q_2(\hat{\Gamma}_Q)_{22}^{-1} Q_2^T(\bar{Z} - \bar{y})(\Gamma_Q^{-1})_{11}^{-1}\}. \qquad \text{(A.7)}$$

Note that the mean of this distribution is totally independent of $\hat{\Gamma}$ and that it is obtained by simply replacing $\hat{\Gamma}$ by Γ in (A.6). Since (A.6) contains all dependence on $\hat{\Gamma}$ in $\hat{\xi}_{\text{EGLS}}$ it follows that

$$E_{\hat{\Gamma}}(\hat{\xi}_{\text{EGLS}} \mid \hat{B}, \bar{Z} - \bar{y}) = \hat{\xi}_{\text{GLS}}. \qquad \text{(A.8)}$$

Now we are prepared for the calculation of $\text{var}(\hat{\xi}_{\text{EGLS}})$. We split the total variance by conditioning on $\hat{\Gamma}$, and use (A.8) to write

$$\begin{aligned}\operatorname{var}(\hat{\xi}_{\mathrm{EGLS}}) &= \operatorname{var}_{\hat{B},\bar{Z}-\bar{y}}\{E_{\hat{\Gamma}}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y})\} + E_{\hat{B},\bar{Z}-\bar{y}}\{\operatorname{var}_{\hat{\Gamma}}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y})\} \\ &= \operatorname{var}(\hat{\xi}_{\mathrm{GLS}}) + E_{\hat{B},\bar{Z}-\bar{y}}\{\operatorname{var}_{\hat{\Gamma}}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y})\} \\ &= \operatorname{var}(\hat{\xi}_{\mathrm{GLS}}) + \text{Term 3}\end{aligned}$$

Hence, relative to $\hat{\xi}_{\mathrm{GLS}}$ we have an additional contribution from the conditional variance over $\hat{\Gamma}$, denoted Term 3.

Term 3

We require the variance over the distribution of $\hat{\Gamma}$, and this is obtained by splitting the variance once more, now by an inner conditioning on the submatrix $(\hat{\Gamma}_Q)_{22}$:

$$\begin{aligned}\operatorname{var}_{\hat{\Gamma}}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y}) = \operatorname{var}_{(\hat{\Gamma}_Q)_{22}}\{E(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y}, (\hat{\Gamma}_Q)_{22}) \mid \hat{B}, \bar{Z}-\bar{y}\} \\ + E_{(\hat{\Gamma}_Q)_{22}}\{\operatorname{var}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y}, (\hat{\Gamma}_Q)_{22}) \mid \hat{B}, \bar{Z}-\bar{y}\}.\end{aligned}$$

By (A.7) the conditional mean in the first term is independent of $\hat{\Gamma}$, so this term vanishes. For the second term, starting from (A.5) we have

$$\begin{aligned}\operatorname{var}_{\hat{\Gamma}}(\hat{\xi}_{\mathrm{EGLS}} \mid \hat{B}, \bar{Z}-\bar{y}) &= \operatorname{var}\{(M^{\mathrm{T}})^{-1}(\hat{\Gamma}_Q)_{12}(\hat{\Gamma}_Q)_{22}^{-1}Q_2^{\mathrm{T}}(\bar{Z}-\bar{y}) \mid \hat{B}, \bar{Z}-\bar{y}\} \\ &= E\{(\bar{Z}-\bar{y})^{\mathrm{T}}Q_2(\hat{\Gamma}_Q)_{22}^{-1}Q_2^{\mathrm{T}}(\bar{Z}-\bar{y})(M^{\mathrm{T}})^{-1}(\Gamma_Q^{-1})_{11}^{-1}M^{-1} \mid \hat{B}, \bar{Z}-\bar{y}\} \\ &= \frac{1}{n-q-2}(\bar{Z}-\bar{y})^{\mathrm{T}}Q_2(\Gamma_Q)_{22}^{-1}Q_2^{\mathrm{T}}(\bar{Z}-\bar{y})(\hat{B}\Gamma^{-1}\hat{B}^{\mathrm{T}})^{-1}. \qquad \text{(A.9)}\end{aligned}$$

For the last equality, in the calculation of the expected value, we used the fact (Arnold, 1981, th. 17.15d) that if W is Wishart $W_f(N, \Sigma)$, then $E(W^{-1}) = \Sigma^{-1}/(N-f-1)$, but also we returned in the latter factors of the formula from the decomposition MQ_1^{T} to the original $\hat{B}$.

Now it is time for the expected value over $(\bar{Z}-\bar{y})$ in formula (A.9), which is a quadratic form in $Q_2^{\mathrm{T}}(\bar{Z}-\bar{y})$. The usual calculation rule for expected values of quadratic forms yields (for fixed Q)

$$\begin{aligned}&E\{(\bar{Z}-\bar{y})^{\mathrm{T}}Q_2(\Gamma_Q)_{22}^{-1}Q_2^{\mathrm{T}}(\bar{Z}-\bar{y}) \mid \hat{B}\} \\ &\quad = E\{(\bar{Z}-\bar{y}) \mid \hat{B}\}^{\mathrm{T}}Q_2(\Gamma_Q)_{22}^{-1}Q_2^{\mathrm{T}}E\{(\bar{Z}-\bar{y}) \mid \hat{B}\} + \operatorname{tr}\operatorname{var}\{(\Gamma_Q)_{22}^{-1/2}Q_2^{\mathrm{T}}(\bar{Z}-\bar{y}) \mid \hat{B}\} \\ &\quad = (\xi-\bar{x})^{\mathrm{T}}BQ_2(\Gamma_Q)_{22}^{-1}Q_2^{\mathrm{T}}B^{\mathrm{T}}(\xi-\bar{x}) + \left(\frac{1}{m}+\frac{1}{n}\right)(q-p). \qquad \text{(A.10)}\end{aligned}$$

The first term of (A.10) appears to have been overlooked by Fujikoshi & Nishii (1986). Its expected value is evaluated approximately by utilizing that $\hat{B}-B \to 0$ in distribution. Note that $\hat{B}Q_2 = MQ_1^{\mathrm{T}}Q_2 = 0$, so we may express the smallness of BQ_2 equivalently in terms of $(\hat{B}-B)$ by substituting $(\hat{B}-B)Q_2$ for each BQ_2 in the first term of the last line of (A.10). As $\hat{B} \to B$, we may assume that the matrices M and (the important) Q in the decomposition of $\hat{B}$ approach deterministic limiting matrices at the same rate. With Q regarded as constant it follows from the matrix normal distribution of $\hat{B}$ that $(\Gamma_Q)_{22}^{-1/2}Q_2^{\mathrm{T}}(\hat{B}-B)^{\mathrm{T}}(\xi-\bar{x})$ is (vector) normal,

$$N_{q-p}\{0, (\xi-\bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi-\bar{x})I_{q-p}\}.$$

Hence, the quadratic form appearing as the first term of (A.10) has the approximate expected value, with a relative error of $O(\delta)$,

$$\operatorname{tr}\operatorname{var}\{(\Gamma_Q)_{22}^{-1/2}Q_2^{\mathrm{T}}(\hat{B}-B)^{\mathrm{T}}(\xi-\bar{x})\} = (q-p)(\xi-\bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi-\bar{x}). \qquad \text{(A.11)}$$

Combining (A.11) with (A.10) and (A.9) we can conclude that

$$\text{Term } 3 \approx \frac{q-p}{n-q-2}\left\{\frac{1}{m}+\frac{1}{n}+(\xi-\bar{x})^{\mathrm{T}}S_{xx}^{-1}(\xi-\bar{x})\right\}(B\Gamma^{-1}B^{\mathrm{T}})^{-1}$$

$$\approx \frac{q-p}{n-q-2}\operatorname{var}(\hat{\xi}_{\mathrm{GLS}}),$$

with the same relative error magnitudes as the corresponding terms of $\operatorname{var}(\hat{\xi}_{\mathrm{GLS}})$. In other words, we have shown that the randomness in $\hat{\Gamma}$ has the effect to increase the variance by a factor

$$1+\frac{q-p}{n-q-2}=\frac{n-p-2}{n-q-2}$$

over the GLS variance. This concludes the proof of theorem 1. □