Received: 19 May 2009, Revised: 17 November 2009, Accepted: 22 November 2009, Published online in Wiley Interscience: 8 January 2010

(www.interscience.wiley.com) DOI: 10.1002/cem.1275

Heteroscedastic controlled calibration model applied to analytical chemistry

Betsabé Blas[a]* and Mônica C. Sandoval[a]

In chemical analyses performed by laboratories, one faces the problem of determining the concentration of a chemical element in a sample. In practice, one deals with the problem using the so-called linear calibration model, which considers that the errors associated with the independent variables are negligible compared with the former variable. In this work, a new linear calibration model is proposed assuming that the independent variables are subject to heteroscedastic measurement errors. A simulation study is carried out in order to verify some properties of the estimators derived for the new model and it is also considered the usual calibration model to compare it with the new approach. Three applications are considered to verify the performance of the new approach. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: linear calibration model; controlled variable; measurement error model; uncertainty; chemical analysis

1. INTRODUCTION

The calibration model, where a response variable is modelled as a function of the independent variable, is defined in two stages. At the first stage (or calibration experiment), a regression model for the relation between the dependent variable Y and the independent variable x is established. Afterwards, at the second stage (or the prediction stage), the model obtained at the previous stage is used to make inference on the unknown value X_0 related to a new observation Y_0.

The usual calibration model [1] (see Appendix B) is commonly used to estimate the concentration X_0 of a chemical species in a test sample. It typically assumes that the independent variable is fixed and it is not subject to error. The estimator given in Reference [1] (Equation (C.3)) is known as a classical estimator. There is another estimator in the literature called inverse estimator which is derived by considering the regression of x on Y. We can find an interesting review in Reference [2] about the wide variety of approaches to both univariate and multivariate calibrations without the assumption of error in the independent variable.

For continuous variables, measurement error is generally characterized as either classical or Berkson type (see Reference [3]). The classical error model is appropriate when an attempt is made to determine x directly, but one is unable to do so because of various errors in measurement. On the other hand, the Berkson-type error [4] arises when the variable X can be set to pre-assigned target values but not achieved exactly. As it was mentioned in References [5], when there exists the Berkson-type case the measurement error is independent of the observed predictor (X) but is dependent on the unobserved true variable x, while on the classical error model, the measurement error is independent of x, but dependent on the observed variable. Due to this difference in the stochastic structure, we have completely different procedures in parameter estimation and inference about the models.

Reference [6] reviews the regression techniques by using the method of least squares with error in both axes and discusses the advantages and limitations of the different approaches considering classical error models.

In applications in analytical chemistry, the independent variable is subject to error which arises from the preparation process of a standard solution. Many studies, such as [7–9], attempt to consider the uncertainties due to the preparation process of the standard solutions by applying the error propagation law to the standard error of the estimator of X_0.

When concentration of the standard solution is pre-fixed by the chemical analyst and a process is carried out attempting to attain it, this process generates errors. Hence, in this case the so-called controlled variable (or Berkson-type variable) [4] arises, where the controlled variable X is defined by the pre-fixed concentration value of the standard solution which is expressed by the equation $X = x + \delta$, where x is the unobserved variable and δ is the measurement error variable.

In Reference [10], the so-called homoscedastic controlled calibration model is proposed. This model is formulated in the framework of the usual calibration model assuming that the independent variable is a controlled variable and the associated measurement errors have *equal* variances.

In References [11] and [12], some methods to compute the uncertainties in certain values obtained through measurements are studied. In Reference [12], the uncertainties of standard solutions are computed and it is observed that these uncertainties depend on the concentration values, so we can observe that the usual calibration model and the homoscedastic controlled calibration model do not seem to be the more suitable ones. This problem motivates us to study a calibration model

* *Correspondence to: B. Blas, Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil*
E-mail: bgblas@yahoo.com.br

a *B. Blas, M. C. Sandoval*
Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil

that considers the errors' variability of the preparation of standard solutions. In this work, we propose a calibration model that incorporates the errors' variability arisen from the preparation process of the standard solution and we call it as *the heteroscedastic controlled calibration model*. This work is a continuation to our previous paper [10] in which the study of the so-called homoscedastic controlled calibration model which assumed *equal* variance errors was undertaken.

The paper is organized as follows. In Section 2, we formulate the heteroscedastic controlled calibration model. In Section 3, a simulation study to test the new approach is presented. In Section 4, three applications are considered which show that the proposed model seems to be more adequate. Section 5 presents our concluding remarks. Finally, some tables showing the results of the simulation study, the usual calibration model and the maximum likelihood estimator via the expectation-maximization (EM) method are presented in Appendices A, B and C, respectively.

2. THE PROPOSED MODEL

Among the relevant problems in chemical analysis is the one related to the estimation of the concentration X_0 of a chemical compound in a given sample. In order to tackle this problem a statistical calibration model is used, which is defined by a two-step process. This problem has been considered in References [13] and [14].

The first stage of the calibration model is given by data points (X, Y) which is determined in an experiment where the independent variable X is the one that the experimenter selects. For instance, the concentrations of the standard solutions that a chemist prepares are independent variables since any concentration may be chosen. The dependent variable Y is a measurable property of the independent variable. For example, the dependent variable may be the amount of intensity supplied by the plasma spectrometry method, since the intensity depends on the concentration.

In the second stage of the calibration model, a suitable sample related to the unknown concentration X_0 is prepared in order to obtain the measurements Y_0.

When the standard concentration X is fixed by the analyst, the process of preparation attempting to get it produces an error δ, and the unobserved quantity attained is x. Considering the usual calibration model defined by Equations (C.1) and (C.2) in Appendix B and the equation $X = x + \delta$, we define the heteroscedastic controlled calibration model as

$$Y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

$$X_i = x_i + \delta_i, \qquad i = 1, 2, \ldots, n, \tag{2}$$

$$Y_{0i} = \alpha + \beta X_0 + \epsilon_i, \qquad i = n+1, n+2, \ldots, n+k. \tag{3}$$

Where n is the number of different observed points (X_i, Y_i), k is a number of replicate measurements of the response corresponding to the unknown X_0.

We consider the following conditions:

- $X_1, X_2, \ldots, X_n$ take pre-fixed values by the analyst.
- $\epsilon_1, \epsilon_2, \ldots, \epsilon_{n+k}$ are independent and normally distributed with mean 0 and variance σ^2_ϵ.
- $\delta_1, \delta_2, \ldots, \delta_n$ are independent and normally distributed with mean 0 and variance $\delta^2_{\sigma_i}$, $i = 1, \ldots, n$ are supposed to be known.
- δ_i, $i = 1, \ldots, n$ and ϵ_i, $i = 1, \ldots, n+k$ are independent.

Observe that in the model described above we only consider the case when the variances $\sigma^2_{\delta_i}$, $i = 1, \ldots, n$ are known. It is a generalization of the homoscedastic controlled calibration model discussed in Reference [10], when it is considered $\sigma^2_{\delta_i} = \sigma^2_\delta$ for all i and the known σ^2_δ case. This new model is also a generalization of the usual calibration model in which one takes $\delta_i = 0, i = 1, \ldots, n$.

In this paper, the approach considered is based on the likelihood function [15]. For the heteroscedastic controlled calibration model, the logarithm of the likelihood function is given by

$$\begin{aligned} l(\alpha, \beta, X_0, \sigma^2_\epsilon) \propto & -\frac{1}{2}\sum_{i=1}^{n}\log(\gamma_i) - \frac{k}{2}\log(\sigma^2_\epsilon) \\ & -\frac{1}{2}\left[\sum_{i=1}^{n}\frac{(Y_i - \alpha - \beta X_i)^2}{\gamma_i} \right. \\ & \left. + \sum_{i=n+1}^{n+k}\frac{(Y_{0i} - \alpha - \beta X_0)^2}{\sigma^2_\epsilon}\right], \end{aligned} \tag{4}$$

where $\gamma_i = \sigma^2_\epsilon + \beta^2\sigma^2_{\delta_i}$, $i = 1, \ldots, n$. Solving $\partial l/\partial\alpha = 0$ and $\partial l/\partial X_0 = 0$ one can get the maximum likelihood estimator of α and X_0 given, respectively, by

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \quad \text{and} \quad \hat{X}_0 = \frac{\bar{Y}_0 - \hat{\alpha}}{\hat{\beta}}. \tag{5}$$

From Equations (4) and (5), it follows that the logarithm of the likelihood function for $(\alpha, \beta, X_0, \sigma^2_\epsilon)$ can be written as

$$\begin{aligned} l(\hat{\alpha}, \beta, \hat{X}_0, \sigma^2_\epsilon) \propto & -\frac{1}{2}\sum_{i=1}^{n}\log(\gamma_i) - \frac{k}{2}\log(\sigma^2_\epsilon) \\ & -\frac{1}{2}\left[\sum_{i=1}^{n}\frac{[(Y_i - \bar{Y}) - \beta(X_i - \bar{X})]^2}{\gamma_i} \right. \\ & \left. + \frac{1}{\sigma^2_\epsilon}\sum_{i=n+1}^{n+k}(Y_{0i} - \bar{Y}_0)^2\right]. \end{aligned} \tag{6}$$

Making $\partial l/\partial\beta = 0$, $\partial l/\partial\sigma^2_\epsilon = 0$ in the logarithm of the likelihood function (6), we have the following equations:

$$\sum_{i=1}^{n}\frac{\beta\sigma^2_{\delta_i}\left[\gamma_i - (Y_i - \alpha - \beta X_i)^2\right]}{\gamma_i^2} = \sum_{i=1}^{n}\frac{X_i(Y_i - \alpha - \beta X_i)}{\gamma_i}, \tag{7}$$

$$\sum_{i=1}^{n}\frac{\gamma_i - (Y_i - \alpha - \beta X_i)^2}{\gamma_i^2} = \sum_{i=n+1}^{n+k}\frac{(Y_{0i} - \bar{Y}_0)^2}{\sigma^4_\epsilon} - \frac{k}{\sigma^2_\epsilon}. \tag{8}$$

The estimates of β and σ^2_ϵ can be obtained through some iterative method that solves Equations (7) and (8).

Another method for finding the estimator of $\boldsymbol{\theta} = (\alpha, \beta, X_0, \sigma^2_\epsilon)$ is the EM algorithm, which is a method for finding maximum-likelihood estimates of population parameters of underlying distribution from a given incomplete data set, originally introduced in Reference [16]. It is a widely applicable approach to the iterative computation of maximum likelihood estimators. In Appendix C we present this method for our new approach, which gives the same estimates than maximizing the likelihood function directly described above.

The Fisher expected information $I(\boldsymbol{\theta}) = E\left[\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}\right]$ is given by

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{\gamma_i} + \frac{k}{\sigma_\epsilon^2} & \sum_{i=1}^{n} \frac{X_i}{\gamma_i} + \frac{kX_0}{\sigma_\epsilon^2} & \frac{k\beta}{\sigma_\epsilon^2} & 0 \\ \sum_{i=1}^{n} \frac{X_i}{\gamma_i} + \frac{kX_0}{\sigma_\epsilon^2} & \sum_{i=1}^{n} \frac{X_i^2}{\gamma_i} + 2\beta^2 \sum_{i=1}^{n} \frac{\sigma_{\delta_i}^4}{\gamma_i^2} + \frac{kX_0^2}{\sigma_\epsilon^2} & \frac{k\beta X_0}{\sigma_\epsilon^2} & \beta \sum_{i=1}^{n} \frac{\sigma_{\delta_i}^2}{\gamma_i^2} \\ \frac{k\beta}{\sigma_\epsilon^2} & \frac{k\beta X_0}{\sigma_\epsilon^2} & \frac{k\beta^2}{\sigma_\epsilon^2} & 0 \\ 0 & \beta \sum_{i=1}^{n} \frac{\sigma_{\delta_i}^2}{\gamma_i^2} & 0 & \sum_{i=1}^{n} \frac{1}{2\gamma_i^2} + \frac{k}{2\sigma_\epsilon^4} \end{pmatrix}.$$

When $k = qn$, $q \in Q^+$ and $n \to \infty$, the estimator $\hat{\boldsymbol{\theta}}$ is approximately normally distributed with mean $\boldsymbol{\theta}$ and variance $I(\boldsymbol{\theta})^{-1}$. Since our main interest is to obtain the estimate of X_0 and its variance estimator, we obtain the approximate variance to order n^{-1} for $\hat{X}_0$ from the Fisher expected information matrix, which is given by

$$V(\hat{X}_0) = \frac{\sigma_\epsilon^2}{\beta^2}\left[\frac{1}{n} + \frac{1}{k} - \frac{E_1}{n\sigma_\epsilon^2 E_2}\right], \tag{9}$$

where E_1 and E_2 are given by

$$\begin{aligned} E_1 = & -n\left(\sigma_\epsilon^4 \sum_{i=1}^{n} \frac{1}{\gamma_i^2} + k\right) \sum_{i=1}^{n} \frac{(X_i - X_0)^2}{\gamma_i} + \sigma_\epsilon^2\left(k + \sigma_\epsilon^4\right) \\ & \times \left(1 + \sum_{i=1}^{n} \frac{1}{\gamma_i^2}\right)\left(\sum_{i=1}^{n} \frac{X_i^2}{\gamma_i} \sum_{i=1}^{n} \frac{1}{\gamma_i} - \left(\sum_{i=1}^{n} \frac{X_i}{\gamma_i}\right)^2\right) \\ & + 2\beta^2\left(\sigma_\epsilon^2 \sum_{i=1}^{n} \frac{1}{\gamma_i} - n\right) \\ & \times \left(\sigma_\epsilon^4\left(\sum_{i=1}^{n} \frac{\sigma_{\delta_i}^4}{\gamma_i^2} \sum_{i=1}^{n} \frac{1}{\gamma_i^2} - \left(\sum_{i=1}^{n} \frac{\sigma_{\delta_i}^2}{\gamma_i^2}\right)^2\right) + k \sum_{i=1}^{n} \frac{\sigma_{\delta_j}^4}{\gamma_i^2}\right), \\ E_2 = & \left(\sigma_\epsilon^4 \sum_{i=1}^{n} \frac{1}{\gamma_i^2} + k\right)\left(\sum_{i=1}^{n} \frac{X_i^2}{\gamma_i} \sum_{i=1}^{n} \frac{1}{\gamma_i} - \left(\sum_{i=1}^{n} \frac{X_i}{\gamma_i}\right)^2\right) \\ & + 2\beta^2\left[\sigma_\epsilon^4 \sum_{i=1}^{n} \frac{1}{\gamma_i}\left(\sum_{i=1}^{n} \frac{\sigma_{\delta_i}^4}{\gamma_i^2} \sum_{i=1}^{n} \frac{1}{\gamma_i^2} - \left(\sum_{i=1}^{n} \frac{\sigma_{\delta_i}^2}{\gamma_i^2}\right)^2\right)\right. \\ & \left. + k \sum_{i=1}^{n} \frac{\sigma_{\delta_j}^4}{\gamma_i^2} \sum_{i=1}^{n} \frac{1}{\gamma_i}\right] \end{aligned}$$

Note that when $\sigma_{\delta_i}^2 = 0,\ i = 1, \dots, n$, expression (9) is reduced to the variance of the usual model given in (C.5) and when $\sigma_{\delta_i}^2 = \sigma_\delta^2$ (for all i) the expression (9) is also reduced to the variance of the homoscedastic model when σ_δ^2 is known (see Equation (2.12) of Reference [10]).

In order to construct a confidence interval for X_0 we consider the interval (C.7), where $\hat{V}(\hat{X}_{0C})$ is the estimated variance that follows from Equation (9).

3. SIMULATION STUDY

We present a simulation study to compare the performance of the estimators obtained from the heteroscedastic controlled calibration model (Proposed-M) with the results obtained by considering the usual model (Usual-M).

Three thousand samples generated from the Proposed-M were considered. In all the samples, the parameters α and β take the values 0.1 and 2, respectively. The range of values for the controlled variable was [0,2]. The fixed values for the controlled variable were $x_1 = 0,\ x_i = x_{i-1} + 2/(n-1),\ i = 2, \dots, n$, and the parameter values X_0 were 0.01 (extreme inferior value), 0.8 (near to the central value) and 1.9 (extreme superior value). For the first and second stages, we consider the sample of sizes $n = 5,\ 20,\ 100,\ 5000$ and $k = 2,\ 20,\ 100,\ 500$, respectively. We consider $\sigma_\epsilon^2 = 0.04$ and the maximum parameter values of σ_δ^2 as $\max\{\sigma_{\delta_i}^2\}_{i=1}^{n} = 0.1$. We consider $\sigma_{\delta_i}^2 = i \times 0.1/n$ for $i = 1, \dots, n$.

The empirical mean bias is given by $\sum_{j=1}^{3000}(\hat{X}_0 - X_0)/3000$ and the empirical mean squared error (MSE) is given by $\sum_{j=1}^{3000}(\hat{X}_0 - X_0)^2/3000$. The mean estimated variance of $\hat{X}_0$ is given by $\sum_{j=1}^{3000}\hat{V}(\hat{X}_0)/3000$. The theoretical variances of $\hat{X}_0$ are referred to expressions (C.5) and (9) evaluated on the relevant parameter values.

In Table A.I (Appendix A), we observe that, in general, the bias of $\hat{X}_0$ from the usual model is smaller than the value supplied by the proposed model, whereas the outcome from the usual model is greater compared with MSE of the proposed model. Also, we observe that the mean estimated variance from the proposed model is closer to the theoretical variance as compared to the outcome from the usual model.

Analysing Table A.II (Appendix A), we observe that the amplitude of the intervals for the parameter X_0 from the proposed model, in most of the cases, is smaller when compared with the amplitude from the usual model. When n is large, the amplitude from the usual model is large; this makes the covering percentage from the usual model to be overestimated (approaching 100%). Since the confidence interval was constructed with a 95% confidence level and, in most of the cases of n, k and X_0 the covering

Table I. Concentration (mg/g), uncertainty($u(X_i)$) and intensity of the standard solutions of chromium, cadmium and lead elements

Chromium element			Cadmium element			Lead element		
X_i	$u(X_i)$	Intensity	X_i	$u(X_i)$	Intensity	X_i	$u(X_i)$	Intensity
0.05	0.00016	6455.900	0.05	0.00016	4.89733	0.05	0.00015	0.9471
0.11	0.00027	13042.933	0.10	0.00027	9.706	0.10	0.00025	1.46833
0.26	0.00040	32621.733	0.25	0.00041	23.41333	0.26	0.00039	3.09033
0.79	0.00122	97364.500	0.73	0.00122	69.73	0.77	0.00117	8.40533
1.05	0.00161	129178.100	1.01	0.00168	96.85667	1.01	0.00155	10.92667

Table II. Intensity of the sample solutions of chromium, cadmium and lead elements

Chromium element	Cadmium element	Lead element
10173.6	5.066	1.303
10516.9	5.027	1.290
10352.2	5.085	1.341

percentage from the proposed model approaches 95%, indicating that the proposed model is more suitable.

4. APPLICATION

In this section, we illustrate the usefulness of the proposed model by applying it to the data supplied by the chemical laboratory of the 'Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT)'–Brazil. The outcome from the proposed approach is also compared with the results from the usual model. Our main interest is to estimate the unknown concentration value X_0 of a sample of the chemical elements such as chromium, cadmium and lead.

Table I presents the fixed values of concentration for the standard solutions with their related uncertainty ($u(X_i)$) and the corresponding intensities for the chromium, cadmium and lead elements. The uncertainties considered are computed using the method recommended by the ISOGUM guide (see Reference [17]) and the intensities are supplied by the plasma spectrometry method. These data are referred to as the first stage of the heteroscedastic controlled calibration model.

Moreover, Table II presents the intensities of the sample solutions of chromium, cadmium and lead elements. These data are referred to as the second stage of the calibration model.

From Table I we verify that the uncertainty values increase with the concentration values.

We consider $\sigma^2_{\delta_i} = u(X_i)^2$. The expanded uncertainty $U(\hat{X}_0)$ is obtained multiplying the squared root of the estimate of variance of $\hat{X}_0$ by the value 1.96 (see References [7] and [12]).

We use the *optim* command from the R-project program to estimate the parameters β and σ^2_ϵ on the likelihood function of the proposed model (6). We use as initial point the estimates from $\hat{\beta} = \sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{Y})/\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $\hat{\sigma}^2_\epsilon = \sum_{i=1}^{n}(Y_{0i} - \bar{Y}_0)^2/n$, which are the estimators from the homoscedastic controlled calibration model when σ^2_δ is unknown [10].

Table III presents estimates of $\alpha,\ \beta,\ X_0,\ V(\hat{X}_0)$ and the expanded uncertainty, $U(\hat{X}_0)$, from the proposed model (Proposed-M) of chromium, cadmium and lead elements. Also, we present the estimates obtained from usual calibration model (Usual-M) to observe the performance of both models.

In Table III, for cadmium and lead elements, we observe that the estimates of α, β and X_0 from the Proposed-M and Usual-M are the same. For the chromium element, there are small differences. Also, we observe that for the chromium element there is a small difference between the estimates of X_0 and $U(\hat{X}_0)$ obtained respectively from the usual model and the proposed model. Despite the relevant estimates of α, β and X_0 from both approaches for cadmium and lead element are similar, the estimates of $V(\hat{X}_0)$ and $U(\hat{X}_0)$ differ considerably, the estimates obtained adopting the usual model is greater than the estimates outcome supplied by the proposed model.

5. CONCLUDING REMARKS

In many applications in chemical analysis the independent variable is measured with Berkson-type error. Due to the fact that the error in the independent variable is of the Berkson type,

Table III. Estimates of $\alpha,\ \beta,\ X_0,\ V(\hat{X}_0)$ and $U(\hat{X}_0)$ related to usual and heteroscedastic model, for the chromium, cadmium and lead element

	Chromium element		Cadmium element		Lead element	
Parameters	Usual-M	Proposed-M	Usual-M	Proposed-M	Usual-M	Proposed-M
α	134.9469	124.2801	0.454801	0.454801	−0.3822126	−0.3822126
β	123003.7	123027.3	10.54381	10.54381	94.29881	94.29881
X_0	0.08302691	0.08309769	0.08123556	0.08123556	0.05770535	0.05770535
$V(\hat{X}_0)$	4.357870e-06	4.474395e-06	7.898643e-05	4.440342e-06	0.0001181068	7.237226e-08
$U(\hat{X}_0)$	0.004091601	0.004145942	0.01741936	0.004130135	0.02130068	0.000527281

we believe that our approach is the suitable one. In analytical chemistry applications, the Usual-M is used assuming that the measurement errors are negligible. However, we have considered the Usual-M in our study just to show that the incorrect use of it gives appreciable differences on the estimates of $V(\hat{X}_0)$ and $U(\hat{X}_0)$ compared with the corresponding estimates from the proposed model. Thus, we have shown that the measurement errors might not be always negligible, so we can conclude that the usual model is not always the appropriate one in this context.

Various aspects of the model studied above deserve attention in future research, e.g. it is not considered the error arisen from the test sample solution preparation and the proposed model can be studied by considering other type of distribution of the errors, such as the skew normal distribution [18]. In particular, one of the drawbacks of the usual model is that it does not consider the error in the independent variable; we believe that despite the fact that this error could be very small, it must be considered as an important property of the calibration model.

Acknowledgements

The authors are grateful to Prof. Dr Heleno Bolfarine for carefully reading the manuscript and Dr Olga Satomi from 'Instituto de Pesquisas Tecnológicas'—IPT for collaboration in a previous work. The authors are indebted to the referees for valuable comments and suggestions. Betsabé Blas has been supported by a Brazilian CNPq grant.

REFERENCES

1. Shukla GK. On the problem of calibration. *Technometrics* 1972; **14**: 547.
2. Osborne C. Statistical calibration: a review. *Int. Stat. Rev.* 1991; **59**: 309.
3. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models*. Chapman & Hall: London, 1995.
4. Berkson J. Are there two regression? *J. Am. Stat. Assoc.* 1950; **45**: 164.
5. Wang L. Estimation of nonlinear Berkson-type measurement error models. *Stat. Sin.* 2003; **13**: 1201.
6. Riu J, Rius FX. Univariate regression models with errors in both axes. *J. Chemom.* 1995; **9**: 343.
7. EURACHEM/CITAC Guide "Quantifying Uncertainty in Analytical Measurement" 2nd edn (2000). Williams A, Ellison SLR, Roesslein M (eds). ISBN: 0 948926 15 5. Available from the Eurachem Secretariat (see http://www.eurachem.org/).
8. Chui QSH, Zucchini RR, Lichtig J. Qualidade de medições em química analítica. Estudo de caso: determinação de cádmio por espectrometria de absorção atômica com chama. *Química Nova* 2001; **24**: 374.
9. Bruggemann L, Wenrich R. Evaluation of measurement uncertainty for analytical procedures using a linear calibration function. *Accred. Qual. Assur.* 2002; **7**: 269.
10. Blas BG, Sandoval MC, Satomi O. Homoscedastic controlled calibration model. *J. Chemom.* 2007; **21**: 145.
11. Ballico M. Limitations of the Welch-Satterthwaite approximation for measurement uncertainty calculations. *Metrologia* 2000; **37**: 61.
12. Blas BG. Calibração controlada aplicada à química analítica. São Paulo: IME-USP. Dissertação de Mestrado, 2005.
13. Tallis GM. Note on a calibration problem. *Biometrika* 1969; **56**: 505.
14. Lwin T, Maritz JS. A note on the problem of statistical calibration. *Appl. Stat.* 1980; **29**: 135.
15. Severini, Thomas A. *Likelihood Methods in Statistics*. Oxford University Press: USA, 2000.
16. Dempster AP, Laird N, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* 1977; **39**: 1.
17. International Organization for Standardization - Guide to the Expression of Uncertainty in Measurement, Geneva, 1993. Revised and reprinted in 1995 (ISO 95).
18. Azzalini A. A class of distributions which includes the normal ones. *Scand. J. Stat.* 1985; **12**: 171.

APPENDIX

A: SIMULATION RESULTS

Table A.I. Empirical bias and mean squared error, the mean estimated variance and theoretical variance of $\hat{X}_0$

X_0	n	k	Usual-M		Proposed-M		Usual-M	Proposed-M	Theoretical variance
			Bias	MSE	Bias	MSE	$\hat{V}(\hat{X}_0)$	$\hat{V}(\hat{X}_0)$	$V(\hat{X}_0)$
0.01	5	2	−0.0156	0.0350	−0.0318	0.0334	0.0398	0.0143	0.0257
		20	−0.0236	0.0319	−0.0445	0.0278	0.0131	0.0156	0.0211
		100	−0.0183	0.0306	−0.0429	0.0276	0.0084	0.0155	0.0207
	20	2	−0.0076	0.0119	−0.0049	0.0100	0.0365	0.0053	0.0097
		20	−0.0055	0.0074	−0.0073	0.0055	0.0081	0.0036	0.0051
		100	−0.0059	0.0068	−0.0101	0.0050	0.0036	0.0033	0.0047
	100	2	0.0003	0.0063	0.0020	0.0059	0.0315	0.0047	0.0059
		20	−0.0023	0.0019	−0.0020	0.0014	0.0046	0.0011	0.0014
		100	−0.0014	0.0015	−0.0013	0.0010	0.0017	0.0007	0.0010
	5000	2	0.0008	0.0055	0.0008	0.0055	0.0300	0.0050	0.0050
		20	0.0000	0.0005	0.0000	0.0005	0.0030	0.0005	0.0005
		100	−0.0003	0.0001	−0.0004	0.0001	0.0006	0.0001	0.0001
		500	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000
0.8	5	2	0.0061	0.0193	0.0025	0.0202	0.0254	0.0089	0.0167
		20	0.0033	0.0135	−0.0019	0.0139	0.0047	0.0081	0.0122
		100	0.0014	0.0132	−0.0037	0.0137	0.0029	0.0078	0.0118
	20	2	0.0015	0.0077	0.0015	0.0077	0.0291	0.0042	0.0074
		20	0.0016	0.0032	0.0009	0.0032	0.0036	0.0020	0.0029
		100	−0.0005	0.0026	−0.0018	0.0026	0.0012	0.0016	0.0025
	100	2	0.0010	0.0055	0.0014	0.0054	0.0299	0.0044	0.0055

(Continues)

Table A.I. (Continued)

X_0	n	k	Usual-M		Proposed-M		Usual-M	Proposed-M	Theoretical variance
			Bias	MSE	Bias	MSE	$\hat{V}(\hat{X}_0)$	$\hat{V}(\hat{X}_0)$	$V(\hat{X}_0)$
		20	0.0006	0.0010	0.0007	0.0010	0.0031	0.0008	0.0010
		100	−0.0001	0.0006	−0.0001	0.0006	0.0007	0.0004	0.0006
	5000	2	0.0014	0.0051	0.0014	0.0051	0.0300	0.0050	0.0050
		20	0.0006	0.0005	0.0006	0.0005	0.0030	0.0005	0.0005
		100	0.0001	0.0001	0.0001	0.0001	0.0006	0.0001	0.0001
		500	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
1.9	5	2	0.0500	0.0802	0.0582	0.0704	0.0432	0.0275	0.0482
		20	0.0314	0.0620	0.0503	0.0562	0.0124	0.0278	0.0435
		100	0.0434	0.0645	0.0594	0.0587	0.0085	0.0283	0.0430
	20	2	0.0070	0.0213	0.0054	0.0185	0.0351	0.0086	0.0166
		20	0.0117	0.0160	0.0127	0.0132	0.0075	0.0067	0.0118
		100	0.0099	0.0161	0.0104	0.0130	0.0033	0.0064	0.0114
	100	2	0.0016	0.0080	0.0007	0.0076	0.0312	0.0055	0.0074
		20	0.0019	0.0035	0.0014	0.0029	0.0043	0.0017	0.0028
		100	0.0001	0.0031	0.0009	0.0025	0.0015	0.0013	0.0024
	5000	2	−0.0008	0.0051	−0.0009	0.0051	0.0300	0.0050	0.0050
		20	−0.0003	0.0006	−0.0004	0.0006	0.0030	0.0005	0.0005
		100	−0.0003	0.0002	−0.0002	0.0001	0.0006	0.0001	0.0001
		500	0.0000	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001

Table A.II. Covering percentage (%) and amplitude (A) of the intervals with a 95% confidence level for the parameter X_0

X_0	n	k	Usual-M		Proposed-M	
			%	A	%	A
0.01	5	2	89	0.35	78	0.22
		20	78	0.21	89	0.24
		100	70	0.17	88	0.24
	20	2	100	0.37	74	0.13
		20	95	0.17	90	0.12
		100	85	0.12	90	0.11
	100	2	100	0.35	84	0.13
		20	100	0.13	91	0.06
		100	96	0.08	90	0.05
	5000	2	100	0.34	94	0.14
		20	100	0.11	95	0.04
		100	100	0.05	94	0.02
		500	100	0.02	94	0.01
0.8	5	2	90	0.28	78	0.18
		20	73	0.13	87	0.17
		100	63	0.10	87	0.17
	20	2	100	0.33	73	0.11
		20	95	0.12	87	0.09
		100	81	0.07	88	0.08
	100	2	100	0.34	86	0.12
		20	100	0.11	91	0.05
		100	97	0.05	89	0.04
	5000	2	100	0.34	95	0.14
		20	100	0.11	95	0.04
		100	100	0.05	95	0.02
		500	100	0.02	93	0.01

(Continues)

Table A.II. (Continued)

X_0	n	k	Usual-M		Proposed-M	
			%	A	%	A
1.9	5	2	78	0.35	81	0.31
		20	59	0.20	86	0.32
		100	51	0.17	87	0.32
	20	2	98	0.36	78	0.17
		20	81	0.17	84	0.16
		100	62	0.11	84	0.16
	100	2	100	0.34	87	0.14
		20	97	0.13	87	0.08
		100	83	0.08	84	0.07
	5000	2	100	0.34	95	0.14
		20	100	0.11	94	0.04
		100	100	0.05	93	0.02
		500	99	0.02	89	0.01

B: USUAL CALIBRATION MODEL

The first and second stage equations of the usual linear calibration model are given, respectively, by

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, 2 \cdots, n, \tag{C.1}$$

$$Y_{0i} = \alpha + \beta X_0 + \epsilon_i, \quad i = n+1, n+2, \ldots, n+k. \tag{C.2}$$

We consider the following assumptions:

- $x_1, x_2, \ldots, x_n$ take fixed values, which are considered as true values.
- $\epsilon_1, \epsilon_2, \ldots, \epsilon_{n+k}$ are independent and normally distributed with mean 0 and variance σ_ϵ^2.

The model parameters are α, β, X_0 and σ_ϵ^2 and the main interest is to estimate the quantity X_0.

The maximum likelihood estimators of the usual calibration model are given by

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}, \qquad \hat{\beta} = \frac{S_{xY}}{S_{xx}}, \qquad \hat{X}_0 = \frac{\bar{Y}_0 - \hat{\alpha}}{\hat{\beta}}, \tag{C.3}$$

$$\sigma_\epsilon^2 = \frac{1}{n+k}\left[\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}x_i)^2 + \sum_{i=n+1}^{n+k}(Y_{0i} - \bar{Y}_0)^2\right], \tag{C.4}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_{xY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y}),$$

$$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \bar{Y}_0 = \frac{1}{n}\sum_{i=n+1}^{n+k} Y_{0i}.$$

The approximation of order n^{-1} for the variance of $\hat{X}_0$ is given by

$$V(\hat{X}_0) = \frac{\sigma_\epsilon^2}{\beta^2}\left[\frac{1}{k} + \frac{1}{n} + \frac{(\bar{x} - X_0)^2}{nS_{xx}}\right]. \tag{C.5}$$

In order to construct a confidence interval for X_0, we consider that

$$\frac{\hat{X}_0 - X_0}{\sqrt{\hat{V}(\hat{X}_0)}} \xrightarrow{D} N(0, 1), \tag{C.6}$$

hence, the approximated confidence interval for X_0 with a confidence level $(1 - \alpha)$ is given by

$$\left(\hat{X}_0 - z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{X}_0)}, \hat{X}_0 + z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{X}_0)}\right), \tag{C.7}$$

where $z_{\frac{\alpha}{2}}$ is the quantile of order $\left(1 - \frac{\alpha}{2}\right)$ of the standard normal distribution.

C: EM ESTIMATOR

The models (1)–(3) can be written as

$$Y_i|x_i \overset{ind}{\sim} N(\alpha + \beta x_i, \sigma_\epsilon^2), \quad i = 1, \ldots, n, \tag{D.1}$$

$$x_i \overset{ind}{\sim} N(X_i, \sigma_{\delta_i}^2), \quad i = 1, \ldots, n, \tag{D.2}$$

$$Y_{0i} \overset{ind}{\sim} N(\alpha + \beta X_0, \sigma_\epsilon^2), \quad i = 1, \ldots, k. \tag{D.3}$$

The complete-data log-likelihood function $l_c(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Y_0}, \mathbf{x})$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)$, $\mathbf{Y}_0 = (Y_{01}, \ldots, Y_{0k})$ and $\mathbf{x} = (x_1, \ldots, x_n)$, is given by

$$\begin{aligned} l_c(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Y}_0, \mathbf{x}) \propto & -\frac{n+k}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{i=1}^{n}\log(\sigma_{\delta_i}^2) \\ & -\frac{1}{2}\left[\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta x_i)^2\right. \\ & \left. + \frac{1}{\sigma_{\delta_i}^2}\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{k}(Y_{0i} - \alpha - \beta X_0)^2\right]. \end{aligned}$$

For the current value $\boldsymbol{\theta}^{(t)}$, the E-step of the EM-type algorithm requires the evaluation of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)}) = E(l_c(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Y}_0, \mathbf{x})|\mathbf{Y}, \mathbf{Y}_0, \hat{\boldsymbol{\theta}}^{(t)})$, where the expectation is taken with respect to the joint conditional distribution of **x** given **Y** and $\mathbf{Y}_0$. Thus, we have that

$$\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)}) = &-\frac{n+k}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{i=1}^{n}\log(\sigma_{\delta_i}^2)\\
&-\frac{1}{2}\Bigg[\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\left\{(Y_i-\alpha)^2+\beta^2(\hat{x_i^2})^{(t)}-2\beta(\hat{x}_i)^{(t)}(Y_{ij}-\alpha)\right\}\\
&+\frac{1}{\sigma_{\delta_i}^2}\sum_{i=1}^{n}\left((\hat{x_i^2})^{(t)}+X_i^2-2(\hat{x}_i)^{(t)}X_i\right)\\
&+\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{k}(Y_{0i}-\alpha-\beta X_0)^2\Bigg],
\end{aligned}$$

where $\hat{x}_i^{(t)} = E[x_i|\mathbf{Y}, \mathbf{Y}_0, \hat{\boldsymbol{\theta}}^{(t)}]$ and $\hat{x}_i{}^{2(t)} = E[x_i^2|\mathbf{Y}, \mathbf{Y}_0, \hat{\boldsymbol{\theta}}^{(t)}]$, $i = 1, \ldots, n$ are given by

$$\hat{x}_i^{(t)} = X_i + \frac{\hat{\beta}^{(t)}\hat{\sigma}_{\delta_i}^{2(t)}}{\hat{\gamma}_i^{(t)}}(Y_i - \hat{\alpha}^{(t)} - \hat{\beta}^{(t)}X_i), \qquad \text{(D.4)}$$

$$\hat{x_i^2}^{(t)} = \frac{\hat{\sigma}_\epsilon^{2(t)}\hat{\sigma}_{\delta_i}^{2(t)}}{\hat{\gamma}_i^{(t)}} + \left(X_i + \frac{\hat{\beta}^{(t)}\hat{\sigma}_{\delta_i}^{2(t)}}{\hat{\gamma}_i^{(t)}}(Y_i - \hat{\alpha}^{(t)} - \hat{\beta}^{(t)}X_i)\right)^2, \qquad \text{(D.5)}$$

with $\hat{\gamma}_i^{(t)} = \hat{\sigma}_\epsilon^{2(t)} + \hat{\beta}^{2(t)}\hat{\sigma}_{\delta_i}^{2(t)}$.

The M-step requires the maximization of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(t)})$ with respect to $\boldsymbol{\theta}^{(t)}$. The closed-form equation for the M-step is given by

$$\begin{aligned}
\hat{X}_0^{(t+1)} &= \frac{\bar{Y}_0 - \hat{\alpha}^{(t)}}{\hat{\beta}^{(t)}},\\
\hat{\beta}^{(t+1)} &= \frac{\sum_{i=1}^{n}\hat{x}_i{}^{(t)}(Y_i - \hat{\alpha}^{(t)})}{\sum_{i=1}^{n}\hat{x_i^2}^{(t)}},\\
\hat{\alpha}^{(t+1)} &= \bar{Y} - \hat{\beta}^{(t)}\hat{\bar{x}}^{(t)},\\
\hat{\sigma}_\epsilon^{2(t+1)} &= \frac{1}{k+n}\Bigg[\sum_{i=1}^{n}\left((Y_i - \hat{\alpha}^{(t)})^2 + \beta^2\hat{x}_i^{2(t)} - 2\hat{\beta}^{(t)}\hat{x}_i^{(t)}(Y_i - \hat{\alpha}^{(t)})\right)\\
&\quad + \sum_{i=1}^{k}\left(Y_{0i} - \hat{\alpha}^{(t)} - \hat{\beta}^{(t)}\hat{X}_0{}^{(t)}\right)^2\Bigg],
\end{aligned}$$

where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\hat{\bar{x}}^{(t)} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i{}^{(t)}$.

The above algorithm is iterated until a suitable convergence rule is satisfied, e.g., $\|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\|$ is sufficiently small.