



HAL
open science

A non-Bayesian predictive approach for statistical calibration

Noslen Hernández, Rolando Biscay, Isneri Talavera

► To cite this version:

Noslen Hernández, Rolando Biscay, Isneri Talavera. A non-Bayesian predictive approach for statistical calibration. *Journal of Statistical Computation and Simulation*, 2012, 82 (4), pp.529-545. <10.1080/00949655.2010.545060>. <hal-04802674>

HAL Id: hal-04802674

<https://hal.science/hal-04802674v1>

Submitted on 25 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A non-Bayesian predictive approach for statistical calibration

Noslen Hernández^{a*}, Rolando J. Biscay^b and Isneri Talavera^a

^a*Advanced Technologies Application Centre, Havana, Cuba;* ^b*Institute of Mathematics, Physics and Cybernetics, Havana, Cuba*

A non-Bayesian predictive approach for statistical calibration is introduced. This is based on particularizing to the calibration setting the general definition of non-Bayesian (or frequentist) predictive probability density proposed by Harris [*Predictive fit for natural exponential families*, *Biometrika* 76 (1989), pp. 675–684]. The new method is elaborated in detail in case of Gaussian linear univariate calibration. Through asymptotic analysis and simulation results with moderate sample size, it is shown that the non-Bayesian predictive estimator of the unknown parameter of interest in calibration (commonly, a substance concentration) favourably compares with previous estimators such as the classical and inverse estimators, especially for extrapolation problems. A further advantage of the non-Bayesian predictive approach is that it provides not only point estimates but also a predictive likelihood function that allows the researcher to explore the plausibility of any possible parameter value, which is also briefly illustrated. Furthermore, the introduced approach offers a general framework that can be applied for calibrating on the basis of any parametric statistical model, so making it potentially useful for nonlinear and non-Gaussian calibration problems.

Keywords: calibration; non-Bayesian predictive density; non-Bayesian predictive likelihood

1. Introduction

Statistical calibration plays a crucial role in many areas of technology such as pharmacology and Chemometrics [1–6]. The calibration problem can be described as follows. A random vector Y (typically, a measurement recorded by a certain device such as a spectrometer) is related to a vector of variables of interest X (e.g., the amount of a certain chemical) according to a statistical model specified by a conditional probability density $f(\cdot; X, \theta)$ that depends on an unknown vector of parameters θ (e.g., the regression coefficients and the variance of the errors in linear calibration models). The density of X may be imposed by the researcher (controlled or design experiments) or given by nature (observational or natural experiments). A sample \mathcal{D} of independent observations $(x_1, y_1), \dots, (x_n, y_n)$ of (X, Y) is available (training sample). Given a new (or future) observation y_0 of Y that corresponds to an unknown value x_0 of X , the problem is to make statistical inferences about x_0 on the basis of the given statistical model, the data \mathcal{D} and y_0 . If x_0 is inside the range of

*Corresponding author. Email: nhernandez@cenatav.co.cu

the training data (x_1, \dots, x_n) , the calibration problem is said to be one of interpolation; otherwise, it is said an extrapolation problem.

In some experiments, several observations y_{01}, \dots, y_{0m} ($m > 1$) of Y that correspond to the same unknown value x_0 of X are available. But in this paper, we will be mainly interested in settings where only one new observation $y_0 = y_{01}$ is recorded (i.e., $m = 1$). Furthermore, we will restrict attention to situations in which X and Y are not vectors but take values in the real line (i.e., univariate calibration), and the values of the variable X are set by the researcher (i.e., controlled experiments). More specifically, it will be assumed that the statistical model $f(\cdot; X, \theta)$ is a Gaussian univariate linear regression model:

$$Y = \alpha + \beta X + \epsilon. \quad (1)$$

Here, the perturbation ϵ is assumed to follow a normal distribution $N(0, \sigma^2)$, and the (row) vector of unknown parameters is $\theta = (\alpha, \beta, \sigma^2)$.

Several approaches have been proposed for this calibration problem, which can be grouped into three main classes:

1.1. Maximum likelihood or classical estimator

The maximum likelihood method to estimate the unknown parameters θ and x_0 leads to what is called the classical estimate \hat{x}_C of x_0 , which was introduced by Eisenhart [7]:

$$\hat{x}_C = \frac{y_0 - \hat{\alpha}}{\hat{\beta}} = \bar{x} + \frac{1}{\hat{\beta}}(y_0 - \bar{y}), \quad (2)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood (i.e., least squares) estimates of the regression coefficients α and β from the data \mathcal{D} , and \bar{x} and \bar{y} are the mean values of x_i and y_i ($i = 1, \dots, n$), respectively. The estimate \hat{x}_C can also be obtained by inverting the fitted regression line $y_0 = \hat{\alpha} + \hat{\beta}x_0$ with respect to x_0 . If several future observations y_{01}, \dots, y_{0m} ($m > 1$) of Y were available then \hat{x}_C is defined by replacing y_0 in expression (2) by the mean value \bar{y}_0 of y_{01}, \dots, y_{0m} .

It is known that the classical estimator is consistent; i.e., \hat{x}_C converges to x_0 when n and m go to infinite. Also, its asymptotic mean square error is the constant function σ^2/β^2 over the whole range of possible values of x_0 [8]. However, this estimator has been greatly criticized because it does not have finite moments, and so its mean square error (MSE) is infinite [9].

1.2. Bayesian estimators

A number of estimators that have been proposed as alternatives to the classical estimator can be justified from a Bayesian perspective.

In particular, this is the case for the inverse estimator \hat{x}_I introduced by Krutchkoff [10,11]. He derived this estimator by considering the linear regression model of X on Y (i.e., the ‘inverse’ regression in comparison with regressing Y on X), say

$$X = \gamma + \delta Y + \eta;$$

computing the least squares estimates $\hat{\gamma}$ and $\hat{\delta}$ of the coefficients γ and δ of this model, and evaluating the fitted regression line at $Y = y_0$ to obtain

$$\hat{x}_I = \hat{\gamma} + \hat{\delta}y_0.$$

This can also be expressed

$$\hat{x}_I = \bar{x} + \frac{\hat{\beta}}{\hat{\beta}^2 + ((n-2)/n)\hat{\sigma}^2/\sigma_x^2}(y_0 - \bar{y}), \quad (3)$$

where

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2, \\ \hat{y}_i &= \hat{\alpha} + \hat{\beta}x_i, \\ \sigma_x^2 &= \frac{1}{n} \sum (x_i - \bar{x})^2. \end{aligned} \quad (4)$$

The inverse estimator can also be derived as a Bayesian estimator under a non-informative prior distribution [12,13]. Alternatively, it may be obtained by minimizing (over certain class of linear estimators) empirical versions of the Bayesian risk, such as the cross-validation criterion in [14] or the compound error in [15].

Several estimators reduce to or are very similar to the inverse estimator when the number m of observations of Y corresponding to the unknown $X = x_0$ is $m = 1$. This is the case for the estimators proposed by Halperin [16] and by Hagwood [17].

Ended, the estimator defined in [16] is given by

$$\hat{x}_H = (1 - R_M)\bar{x} + R_M\hat{x}_C, \quad (5)$$

where

$$R_M = \frac{\hat{\beta}^2}{((n-2)/nM)\hat{\sigma}^2/\sigma_x^2}$$

and M is some specified function of m . The usual choice $M = m$ leads to $\hat{x}_H = \hat{x}_I$ when $m = 1$, as can be straightforwardly shown by comparing expressions (5) and (3).

Hagwood [17] thought of the calibration problem as an ill-posed inverse problem, the solution on the basis of quadratic regularization of which yields the family of estimators

$$\hat{x}_\lambda = \bar{x} + \frac{\hat{\beta}}{\hat{\beta}^2 + \lambda}(y_0 - \bar{y}),$$

depending on a regularization parameter $\lambda > 0$. He proved that the asymptotic expected cross validation function is minimized by choosing $\lambda = \hat{\lambda}_{cv} = \hat{\sigma}^2/\sigma_x^2$, which results in the estimator

$$\hat{x}_{\hat{\lambda}_{cv}} = \bar{x} + \frac{\hat{\beta}}{\hat{\beta}^2 + \hat{\sigma}^2/\sigma_x^2}(y_0 - \bar{y}).$$

Note that this estimate is practically indistinguishable from \hat{x}_I (compare with Equation (3)), specially when n is moderate or large.

A number of theoretical and simulation studies have been published that compare the classical with inverse estimators in regard to several criteria [8–11,15,18–23]. Their main results can be summarized as follows:

- (i) In comparison with the inverse estimator, the classical estimator shows better performance for extrapolation and worse performance for interpolation both for moderate and large sample sizes. In particular, \hat{x}_I has unbounded asymptotic mean-square error as a function of x_0

(specifically, it is an increasing function of $(x_0 - \bar{x})^2$) while \hat{x}_C has a constant asymptotic mean-square error. Furthermore, the minimum of the asymptotic mean square error function of \hat{x}_1 , which is achieved at $x_0 = \bar{x}$, is less than that (constant) of \hat{x}_C .

- (ii) \hat{x}_C is a consistent estimator while \hat{x}_1 is not.
- (iii) In contrast with \hat{x}_C , \hat{x}_1 has finite moments, and so finite MSE.
- (iv) The finite sample performance of \hat{x}_C deteriorates more drastically than that of \hat{x}_1 when β becomes small and σ^2 large.

1.3. Asymptotic corrections to the classical estimator

A number of corrections or modifications to the classical estimator have been reported in the literature.

Some of them consist of weighted averages of estimators. For example, on the basis of large sample asymptotic analysis, Ali and Singh [24] proposed the following weighted average of the classical estimator \hat{x}_C and \bar{x} :

$$\hat{x}_{AS} = \lambda \hat{x}_C + (1 - \lambda) \bar{x},$$

where

$$\hat{\lambda} = \frac{\hat{\beta}^2 \hat{\delta}^2}{\hat{\beta}^2 \hat{\delta}^2 + \hat{\sigma}^2}$$

and $\hat{\delta} = \hat{x}_C - \bar{x}$ or $\hat{\delta} = \hat{x}_1 - \bar{x}$. The underlying idea is to improve the interpolation performance of the classical estimator \hat{x}_C by shrinking it towards \bar{x} when x_0 is close to \bar{x} . Such a shrinking results to be small when x_0 is far away from \bar{x} , and so it is to be expected that \hat{x}_{AS} retains the good performance of \hat{x}_C for extrapolation.

Another weighted average estimator was introduced by Srivastava and Singh [25] on the basis of small disturbance asymptotic theory:

$$\hat{x}_{SS} = (1 - \lambda) \hat{x}_C + \lambda \hat{x}_1, \quad (6)$$

where $\lambda = 1/(n - 2)$. This estimator can be thought of as a slight correction to the inverse estimator \hat{x}_1 when the sample size n is small, and a slight correction to the classical estimator \hat{x}_C when n is large.

In order to ensure asymptotic unbiasedness, another correction to the classical estimator has been carried out by Naszódi [26], which leads to the estimator:

$$\hat{x}_N = \bar{x} + \frac{\hat{\beta}}{\hat{\beta}^2 + (1/n)\hat{\sigma}^2/\sigma_x^2} (y_0 - \bar{y}). \quad (7)$$

In contrast to the methods just reviewed, the purpose of this paper is to introduce an approach for statistical calibration that is not based on maximum likelihood, Bayesian or asymptotic considerations but on a non-Bayesian predictive framework. For this, the non-Bayesian predictive density introduced by Harris [27] is applied to the calibration setting, providing a predictive density $f_P(\cdot; x_0, \mathcal{D})$ for an observation y corresponding to $X = x_0$. From this, a non-Bayesian predictive estimator \hat{x}_P of x_0 is defined. This alternative approach has a number of appealing features worth studying:

- (i) Likewise the classical estimator and in contrast to Bayesian estimators, it does not require specification of some prior distribution for model parameters, neither the assumption that the design points x_1, \dots, x_n nor the new point x_0 is generated by a common probability density.

The latter makes this approach potentially useful for situations in which the data x_i comes from controlled experiments and the calibration problem could involve extrapolation.

- (ii) Similarly to Bayesian methods and contrary to the classical estimator, it takes into consideration (through the construction of the predictive density f_P) the statistical uncertainty about the model parameters α , β and σ^2 . It is expected that this has a regularizing effect, and so the accuracy of \hat{x}_P does not deteriorate as much as that of \hat{x}_C when the sample size is not large and β becomes small and σ^2 large.
- (iii) It allows one not only to obtain a point estimate \hat{x}_P of x_0 , but also to explore the (predictive) likelihood of any possible value of x_0 .

The paper is organized as follows. In Section 2, the non-Bayesian predictive likelihood and the associated non-Bayesian predictive estimator are defined. This estimator results to be consistent, asymptotically unbiased and has, to first-order expansion, less (asymptotic) MSE than the classical estimator. In Section 3, a simulation study is carried out to compare the performance of the introduced predictive estimator with that of the aforementioned estimators for moderate sample sizes. The feasibility and usefulness of likelihood assessment by means of the (non-Bayesian) predictive likelihood function is also briefly illustrated. Some concluding remarks are given in Section 4.

2. Non-Bayesian predictive calibration

2.1. Non-Bayesian predictive likelihood and predictive estimator

Let \mathcal{Y} be a sample of independent and identically distributed observations y_1, \dots, y_n the density $f(\cdot; \theta_0)$ of which belongs to a family of densities $f(\cdot; \theta)$, which depends on an unknown vector of parameters $\theta \in \Theta$. On the basis of the data \mathcal{Y} , Harris [27] defined the following non-Bayesian predictive density for a future observation Y from the same distribution $f(\cdot; \theta_0)$:

$$f_P(y; \mathcal{Y}) = \int f(y; \theta) \hat{f}_{\hat{\theta}}(\theta) d\theta,$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ_0 on the basis of (y_1, \dots, y_n) , and $\hat{f}_{\hat{\theta}}$ is an estimate of the density $f_{\hat{\theta}}(\cdot; \theta_0)$ of $\hat{\theta}$. Usually, $\hat{f}_{\hat{\theta}}$ is a plug-in estimate resulting from replacing θ_0 by $\hat{\theta}$ in the analytical expression of the density of $\hat{\theta}$; i.e., $\hat{f}_{\hat{\theta}}(\theta) = f_{\hat{\theta}}(\cdot; \hat{\theta})$. This definition may be generalized to regression models. Specifically, for $i = 1, \dots, n$, let y_i be an observation from a density $f(\cdot; x_i, \theta_0)$ that depends on the unknown parameter θ_0 and also on the value x_i of some covariable X , and denote by \mathcal{D} the (past or training) data $(x_1, y_1), \dots, (x_n, y_n)$. A non-Bayesian predictive density for an (future) observation Y from the density $f(\cdot; x_0, \theta_0)$ corresponding to the covariable value $X = x_0$ can be defined as

$$f_P(y; x_0, \mathcal{D}) = \int f(y; x_0, \theta) \hat{f}_{\hat{\theta}}(\theta) d\theta, \quad (8)$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ_0 based on the data \mathcal{D} , and $\hat{f}_{\hat{\theta}}$ is an estimate of the density $f_{\hat{\theta}}(\cdot; \theta_0)$ of $\hat{\theta}$.

This concept will be here applied to construct a non-Bayesian predictive density for the calibration problem (1). For this, it is convenient to rewrite this model in vector form

$$Y = \mathbf{u}\mathbf{y}^\top + \epsilon,$$

where $\mathbf{u} = (1, X)$ and $\mathbf{y} = (\alpha, \beta)$ are row vectors and \top denotes the transpose operator. The vector of parameters of this model is $\theta = (\mathbf{y}, \sigma^2)$. The true value of the vector parameter is denoted by

$\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0, \sigma_0^2)$, where $\boldsymbol{\gamma}_0 = (\alpha_0, \beta_0)$. We will use the estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}, \hat{\sigma}^2)$, where $\hat{\boldsymbol{\gamma}} = (\hat{\alpha}, \hat{\beta})$ is the maximum likelihood (i.e., least squares) estimate of the model coefficients α and β , and $\hat{\sigma}^2$ is the unbiased estimate given by Equation (4). It is known that the density of $\hat{\boldsymbol{\theta}}$ is

$$f_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta}; \boldsymbol{\theta}_0) = f_{(\hat{\boldsymbol{\gamma}}, \hat{\sigma}^2)}(\boldsymbol{\gamma}, \sigma^2; \boldsymbol{\gamma}_0, \sigma_0^2) = f_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma}; \boldsymbol{\gamma}_0, \sigma_0^2) f_{\hat{\sigma}^2}(\sigma^2; \sigma_0^2),$$

where the densities of $\hat{\boldsymbol{\gamma}}$ and $\hat{\sigma}^2$ are, respectively:

$$f_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma}; \boldsymbol{\gamma}_0, \sigma_0^2) = f_{N(\boldsymbol{\gamma}_0, (\mathbf{U}^\top \mathbf{U})^{-1} \sigma_0^2)}(\boldsymbol{\gamma}),$$

$$f_{\hat{\sigma}^2}(v; \sigma_0^2) = \frac{l}{\sigma_0^2} f_{\chi_{(l)}^2} \left(\frac{l}{\sigma_0^2} v \right).$$

Here, $f_{N(\boldsymbol{\gamma}_0, (\mathbf{U}^\top \mathbf{U})^{-1} \sigma_0^2)}$ is the (bivariate) Gaussian density with mean $\boldsymbol{\gamma}_0$ and covariance matrix $(\mathbf{U}^\top \mathbf{U})^{-1} \sigma_0^2$; \mathbf{U} is the $n \times 2$ matrix with rows $\mathbf{u}_i = (1, x_i)$, $i = 1, \dots, n$; and $f_{\chi_{(l)}^2}$ is the chi-square density with $l = n - 2$ degrees of freedom. Therefore, a plug-in estimate of $f_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$ is given by

$$\hat{f}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = \hat{f}_{(\hat{\boldsymbol{\gamma}}, \hat{\sigma}^2)}(\boldsymbol{\gamma}, \sigma^2; \hat{\boldsymbol{\gamma}}, \hat{\sigma}^2) = f_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma}; \hat{\boldsymbol{\gamma}}, \hat{\sigma}^2) f_{\hat{\sigma}^2}(\sigma^2; \hat{\sigma}^2).$$

Taking into consideration that $f(y; x_0, \boldsymbol{\theta}) = f_{N(\mathbf{u}_0 \boldsymbol{\gamma}^\top, \sigma^2)}(y)$, where $\mathbf{u}_0 = (1, x_0)$, and

$$\int f_{N(\mathbf{u}_0 \boldsymbol{\gamma}^\top, \sigma^2)}(y) f_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\gamma}; \hat{\boldsymbol{\gamma}}, \hat{\sigma}^2) d\boldsymbol{\gamma} = f_{N(\mathbf{u}_0 \hat{\boldsymbol{\gamma}}^\top, \mathbf{u}_0 (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{u}_0^\top \hat{\sigma}^2 + \sigma^2)}(y),$$

the predictive density (8) reduces to

$$f_P(y; x_0, \mathcal{D}) = \int_0^\infty f_{N(\mathbf{u}_0 \hat{\boldsymbol{\gamma}}^\top, \mathbf{u}_0 (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{u}_0^\top \hat{\sigma}^2 + v)}(y) f_{\hat{\sigma}^2}(v; \hat{\sigma}^2) dv. \quad (9)$$

In the calibration setting, a new value y_0 of the response variable Y is observed that corresponds to an unknown value x_0 of the covariable X , and the problem of interest is to estimate x_0 .

The likelihoods of different possible values of x_0 may be assessed through what we will call the predictive likelihood function for x_0 . This is defined as the function $L_P(x_0) = f_P(y_0; x_0, \mathcal{D})$ over the range of possible values of x_0 , for the fixed, given data y_0 and \mathcal{D} .

Furthermore, we define the non-Bayesian predictive estimator \hat{x}_P of x_0 by maximizing the predictive likelihood function, that is,

$$\hat{x}_P = \arg \max_{x \in \mathbb{R}} L_P(x). \quad (10)$$

Note that these definitions can be extended to the case in which several observations y_{01}, \dots, y_{0m} ($m > 1$) of Y_0 are available. Specifically, in such a setting the predictive density for the mean \bar{y}_0 of the (future) observations y_{01}, \dots, y_{0m} is given by

$$L_P(x_0) = f_P(\bar{y}_0; x_0, \mathcal{D}) = \int_0^\infty f_{N(\mathbf{u}_0 \hat{\boldsymbol{\gamma}}^\top, \mathbf{u}_0 (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{u}_0^\top \hat{\sigma}^2 + v/m)}(\bar{y}_0) f_{\hat{\sigma}^2}(v; \hat{\sigma}^2) dv, \quad (11)$$

and the predictive estimator is obtained through maximization of this function with respect to x_0 .

2.2. Some asymptotic properties

Asymptotic Taylor expansions to order $1/n$ (see Equations (A5)–(A8) in the Appendixes 1 and 2) lead to the approximations:

$$\hat{x}_P \approx \bar{x} + (1 - \hat{\eta}_n^2) \frac{1}{\hat{\beta}} (y_0 - \bar{y}), \quad (12)$$

$$E(\hat{x}_p) \approx x_0, \quad (13)$$

$$V(\hat{x}_p) \approx \frac{\sigma^2}{\beta^2} + \eta_n^2 \left((x_0 - \bar{x})^2 - \frac{\sigma^2}{\beta^2} + \sigma_x^2 \right), \quad (14)$$

$$\text{MSE}(\hat{x}_p) = E(\hat{x}_p - x_0)^2 \approx \frac{\sigma^2}{\beta^2} + \eta_n^2 \left((x_0 - \bar{x})^2 - \frac{\sigma^2}{\beta^2} + \sigma_x^2 \right), \quad (15)$$

where

$$\hat{\eta}_n^2 = \frac{\hat{\sigma}^2}{\sigma_x^2 \hat{\beta}^2 n},$$

$$\eta_n^2 = \frac{\sigma^2}{\sigma_x^2 \beta^2 n}.$$

Here, the symbol ' \approx ' expresses that the right-hand sides of the equations approximate their left-hand sides except for remainders r_n terms of order $1/n_2$; more specifically, $r_n = O_p(1/n^2)$ in Equation (12) and $r_n = O(1/n^2)$ in Equations (13)–(15) (technical details about Taylor expansions in probability may be found, e.g., in [28]).

A number of asymptotic properties of the predictive estimator \hat{x}_p that follows from this approximations are worth noting:

- (i) Retaining terms to order zero in expansions with respect to $1/n$, the predictive estimator \hat{x}_p coincides with the classical estimator \hat{x}_c (i.e., disregarding the factor $\hat{\eta}_n^2 = O(1/n)$ in Equation (12) one obtains Equation (2)), and so they have the same asymptotic properties to zero-order expansion. Hence, to zero-order expansions it holds that: (a) the MSE error of \hat{x}_p is σ^2/β^2 for all values of x_0 ; and (b) using the MSE as criterion, \hat{x}_p (likewise the classical estimator) is worse than the inverse estimator \hat{x}_i for interpolation (i.e., when x_0 is close enough to \bar{x}) but \hat{x}_p outperforms \hat{x}_i for extrapolation (i.e., when x_0 sufficiently deviates from \bar{x}).
- (ii) It is known (see, e.g., [18]) that, to first-order Taylor expansion, the MSE of the classical estimator is given by

$$\text{MSE}(\hat{x}_c) \approx \frac{\sigma^2}{\beta^2} + \eta_n^2 \left[(x_0 - \bar{x})^2 + 3 \frac{\sigma^2}{\beta^2} + \sigma_x^2 \right]. \quad (16)$$

Comparing Equations (15) and (16) it follows that, to first-order expansions,

$$\text{MSE}(\hat{x}_c) - \text{MSE}(\hat{x}_p) \approx 4 \frac{\sigma^4}{\beta^4 \sigma_x^2 n} > 0$$

for any value of x_0 . Thus, from the viewpoint of MSE to first-order expansions, the introduced estimator \hat{x}_p is better than the classical estimator \hat{x}_c over the whole range of possible values of x_0 . Also note that $\text{MSE}(\hat{x}_c) - \text{MSE}(\hat{x}_p)$ increases when σ^4 becomes large and β^4 and σ_x^2 are small.

- (iii) It is known (see, e.g., [18]) that, to first-order Taylor expansions, the MSE of \hat{x}_i is given by

$$\text{MSE}(\hat{x}_i) \approx \frac{\sigma^2}{\beta^2(1+\Psi)^2} \left[\frac{1}{n} + 1 + \frac{(x_0 - \bar{x})^2}{(n-1)\sigma_x^2} + \frac{\sigma^2(\Psi^2 + 5)}{\beta^2(1+\Psi)^2(n-1)\sigma_x^2} + \frac{\sigma^2(x_0 - \bar{x})^2}{\beta^2\sigma_x^4} \left(1 - \frac{6}{n(1+\Psi)^2} \right) \right], \quad (17)$$

where

$$\Psi = \frac{\sigma^2}{\beta^2 \sigma_x^2}.$$

Comparing expressions (15) and (17) it can be shown that, to first-order expansion, $\text{MSE}(\hat{x}_1) < \text{MSE}(\hat{x}_p)$ when $x_0 = \bar{x}$. This means that the inverse estimator outperforms \hat{x}_p for interpolation, i.e. when x_0 is around \bar{x} . It can also be shown that the difference $\text{MSE}(\hat{x}_p) - \text{MSE}(\hat{x}_1)$ evaluated at $x_0 = \bar{x}$ is an increasing function of Ψ (when $n > 2$). This means that for x_0 in a neighbourhood of \bar{x} the performance of \hat{x}_p compares with \hat{x}_1 deteriorates more notably when σ^2 is large and β^2 and σ_x^2 are small.

On the other hand, it can be shown that for $(x_0 - \bar{x})^2$ large enough it holds that $\text{MSE}(\hat{x}_1) > \text{MSE}(\hat{x}_p)$. That is, for extrapolation problems \hat{x}_p outperforms \hat{x}_1 .

- (iv) In the case in which several observations y_{01}, \dots, y_{0m} ($m > 1$) of Y_0 are available it follows from Equation (11) that, to zero-order expansion, $\hat{x}_p \approx \bar{x} + (\bar{y}_0 - \bar{y})/\hat{\beta}$. This implies that \hat{x}_p is a consistent estimator of x_0 , i.e., \hat{x}_p goes to x_0 when both n and m go to infinite.

Summarizing, to first-order expansions, we can conclude the following concerning the introduced estimator \hat{x}_p : for extrapolation it outperforms not only the inverse estimator but also the classical estimator, while for interpolation its behaviour, though not better than that of the inverse estimator, yields an improvement regarding the classical estimator. Furthermore, the superiority of the introduced estimator over the classical one becomes greater in situations in which σ^2 is large and β^2 small, i.e. when statistical regularization is more advantageous in inverting the regression model to obtain x_0 .

The non-asymptotic behaviour of the introduced non-Bayesian predictive estimator is explored by means of computer simulations in the next section.

3. Simulation study

The feasibility and performance of the introduced non-Bayesian predictive approach for calibration is explored in this section through a simulation study for moderate sample sizes.

Note that the computation of the predictive estimator \hat{x}_p requires to maximize the predictive density $f_P(y_0; x, \mathcal{D})$ with respect to x , and this involves integrals with respect to the densities of $\hat{\beta}$ and $\hat{\sigma}^2$ (see Equation (9)). In practice, the latter integral has no appreciable numerical effect in comparison with simply evaluating at $\sigma = \hat{\sigma}$, i.e., considering $f_{\hat{\sigma}^2}(\sigma^2; \hat{\sigma}^2)$ as a Dirac delta function centred at $\sigma = \hat{\sigma}$. For simplicity, it will be done so in the computations of this simulation study.

The predictive estimator \hat{x}_p will be compared with other estimators reported in the literature in regards to several measures of accuracy. In addition, the usefulness of the introduced approach to assess predictive likelihoods of different possible values of the parameter of interest x_0 will be illustrated.

Likewise the classical estimator, the new approach is mainly oriented to be used for extrapolation problems of calibration (see comments in Sections 1 and 2). In order for an estimator to be considered suitable for extrapolation, it seems natural to impose the condition that, to zero-order approximation, its mean-square error be a bounded function of x_0 over the whole real line. Of all the estimators reviewed in Section 1, only \hat{x}_C , \hat{x}_{SS} and \hat{x}_N fulfilled this property, and in addition they are consistent. They are the focus of this study. Taking into consideration its very common use, the inverse estimator \hat{x}_1 is also included.

For comparing these estimators, sets of data were generated under a number of settings with different values of the sample size n and the parameters α , β and σ^2 in model (1), as well as

Table 1. Values of the model coefficient β , the coefficient of variation CV, the noise to signal ratio NSR and σ/β for the four simulation settings $M1$, $M2$, $M3$ and $M4$ in the simulation study.

	$M1$	$M2$	$M3$	$M4$
β	1.63	0.54	0.09	0.03
CV	0.009	0.029	0.177	0.531
NSR	0.044	0.142	0.867	2.601
σ/β	0.061	0.185	1.111	3.333

different experimental designs for choosing the points x_1, \dots, x_n in the training data \mathcal{D} . Results from some of these settings or scenarios are shown here.

Specifically, we set $\alpha = 0$ and $\sigma = 0.1$ in model (1). Note that, among finite sample sizes not covered by asymptotic analysis (say, $n < 50$), a moderate sample size (n around 25) is notably interesting for many applications in which small samples (say, $n < 10$) are seldom analysed by the researcher. We fix sample size $n = 24$ in this study. The design points $\mathcal{X} = \{x_1, \dots, x_n\}$ were obtained by sampling from a normal distribution with mean 15 and standard deviation 1.5. The actual range of these points results to be $[12.9, 17.5]$, which is so regarded the calibration range. Four settings or models $M1, M2, M3$ and $M4$ were considered that correspond to four different values of β in the model (1) (see Table 1). They are associated with increasing values of the coefficient of variation:

$$\text{CV} = \frac{\sqrt{\text{var}(\hat{\beta})}}{|\beta|} = \frac{\sigma}{\sqrt{n}\sigma_x|\beta|}.$$

Note that CV can be interpreted in terms of a ‘noise to signal ratio’ $\text{NSR} = \sigma/(\sigma_x|\beta|)$, and it is equivalent to σ^2/β^2 for fixed values of n and σ_x . This quantity plays a crucial role in the accuracy of estimates both in asymptotic sense (see Section 2 and Appendix) and finite sample size performance. No matter the type of estimator used for calibration, settings with high values of CV and NSR yield less accurate estimates of x_0 , and will be referred to as ‘bad’ settings in comparison with ‘nice’ settings characterized by low values of CV and NSR. As can be observed in Table 1 the CV and NSR values in the settings of these simulations vary from very nice scenarios (CV $\approx 1\%$, NSR $\approx 4\%$ in $M1$) to very bad scenarios (CV $\approx 50\%$, NSR $\approx 200\%$ in $M4$).

Several criteria are used to evaluate the performance of an estimator \hat{x} of x_0 . The first one is the root mean square error $\text{RMSE} = \sqrt{E(\hat{x} - x_0)^2}$. This criterion cannot be applied to estimators with non-finite second-order moments, such as \hat{x}_C and \hat{x}_{SS} [29]. Thus, we used this criterion only to compare with the estimators \hat{x}_N and \hat{x}_I (see Appendix 1 for a proof of finiteness of second moments of \hat{x}_P). Another criterion is used in order to include all the estimators \hat{x}_P , \hat{x}_C , \hat{x}_{SS} and \hat{x}_N in the assessment. Specifically, for any estimator \hat{x} of x_0 , the 50%–percentile of the distribution of $|\hat{x} - x_0|$ is also adopted as a measure of estimation accuracy, which is denoted by $D50$. To further compare estimators that show similar performance regarding RMSE and $D50$, the Kullback–Leibler (KL) [30] divergence will be used. The KL divergence of an estimate \hat{f} of a density f is defined by

$$\text{KL}(f, \hat{f}) = \int f(y) \ln f(y) dy - \int f(y) \ln \hat{f}(y) dy.$$

For computing and plotting the curves RMSE and $D50$ as functions of x_0 we considered a discrete set of possible x_0 values, namely

$$\mathcal{X}_0 = \{0; 2; 4; 6; 8; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 22; 24; 26; 28; 30\}.$$

Note that \mathcal{X}_0 overpasses the calibration range [12.9, 17.5], hence allowing one to assess the calibration methods not only for interpolation (i.e., within the calibration range) but also for extrapolation (i.e., outside the calibration range).

For each estimation method and each setting or scenarios $M1$, $M2$, $M3$ and $M4$, the following computational steps were carried out:

- (i) For each repetition $b = 1, \dots, B$ (where $B = 2000$):
 - (1a) (Generation of training data) An independent data set $\mathcal{Y}^b = (y_1^b, \dots, y_n^b)$ was simulated, where y_i^b is generated according to the regression model (1) with the parameter values of the specific setting and $X = x_i$, $i = 1, \dots, n$. In this way, the training data $\mathcal{D}^b = \{(x_1, y_1^b), \dots, (x_n, y_n^b)\}$ is obtained.
 - (1b) (Generation of future observation) For each $x_0 \in \mathcal{X}_0$ a new independent observation y_0^b is generated according to the model (1) with the parameter values of the specific setting but assigning $X = x_0$.
 - (1c) (Estimation) For each $x_0 \in \mathcal{X}_0$, the training data \mathcal{D}^b and the new observation y_0^b are used to obtain an estimate \hat{x}_0^b of x_0 , as well as the associated deviation $\hat{x}_0^b - x_0$.
- (ii) (Computation of the curves RMSE and $D50$) The sample of deviations $\hat{x}_0^1 - x_0, \dots, \hat{x}_0^B - x_0$ computed in step (1c) are used to estimate the values of the measures of precision $\text{RMSE}(x_0)$ and $D50(x_0)$ at each $x_0 \in \mathcal{X}_0$.

The resulting curves of RMSE and $D50$ for the different calibration methods are plotted in Figure 1 (for settings $M1$ and $M2$) and Figure 2 (for settings $M3$ and $M4$). They point up a number of properties worth of noting (which were also confirmed by further simulations, not shown here, with a larger variety of settings covering other number of repetitions B , sample sizes n , sets of design points x_i , and values of model parameters α , β and σ). They may be summarized as follows:

- (i) According to both RMSE and $D50$ criteria, all the estimators have worse performances as x_0 goes away from the centre of the calibration range.
- (ii) As can be observed in the upper row of Figure 1, in very nice settings such as $M1$ it is hardly noticeable any difference in performance between the alternative calibration estimators \hat{x}_1 , \hat{x}_C , \hat{x}_{SS} , \hat{x}_N and \hat{x}_P , and all of them show high accuracy.
- (iii) As the setting becomes worse, differences between the estimators begin to be noticeable (see the lower row of Figures 1 and 2). More specifically, it is observed that:
 - (a) Differences in performance between the estimators \hat{x}_P , \hat{x}_C , \hat{x}_{SS} and \hat{x}_N are small in comparison with the difference between any of them and the inverse estimator \hat{x}_1 (see Figure 1 and the upper row of Figure 2) except for very bad settings such as $M4$ (see the lower row of Figure 2).
 - (b) In settings in which any difference in performance between \hat{x}_1 and the group of estimators \hat{x}_P , \hat{x}_C , \hat{x}_{SS} , \hat{x}_N is noticeable, \hat{x}_1 slightly outperforms \hat{x}_P , \hat{x}_C , \hat{x}_{SS} and \hat{x}_N for x_0 values close the centre of the calibration range (i.e., for interpolation) while all the estimators of this latter group greatly outperforms \hat{x}_1 for x_0 values outside or near the limits of the calibration range (i.e., for extrapolation).
 - (c) As the setting becomes quite bad such as $M4$ (see the lower row of Figure 2) some differences in performance between the estimators \hat{x}_P , \hat{x}_C , \hat{x}_{SS} and \hat{x}_N become apparent. The curves of $D50$ (see the right lowest corner of Figure 2) show that \hat{x}_P and \hat{x}_N have similar behaviour and outperform \hat{x}_C and \hat{x}_{SS} , the two latter having similar performance. Thus, the following ordering holds regarding extrapolation: \hat{x}_P and \hat{x}_N are better than \hat{x}_C and \hat{x}_{SS} , which in turn are better than \hat{x}_1 ; while for interpolation: \hat{x}_1 is better than \hat{x}_P and \hat{x}_N , which in turn are better than \hat{x}_C and \hat{x}_{SS} . Summarizing, \hat{x}_P and \hat{x}_N are the best

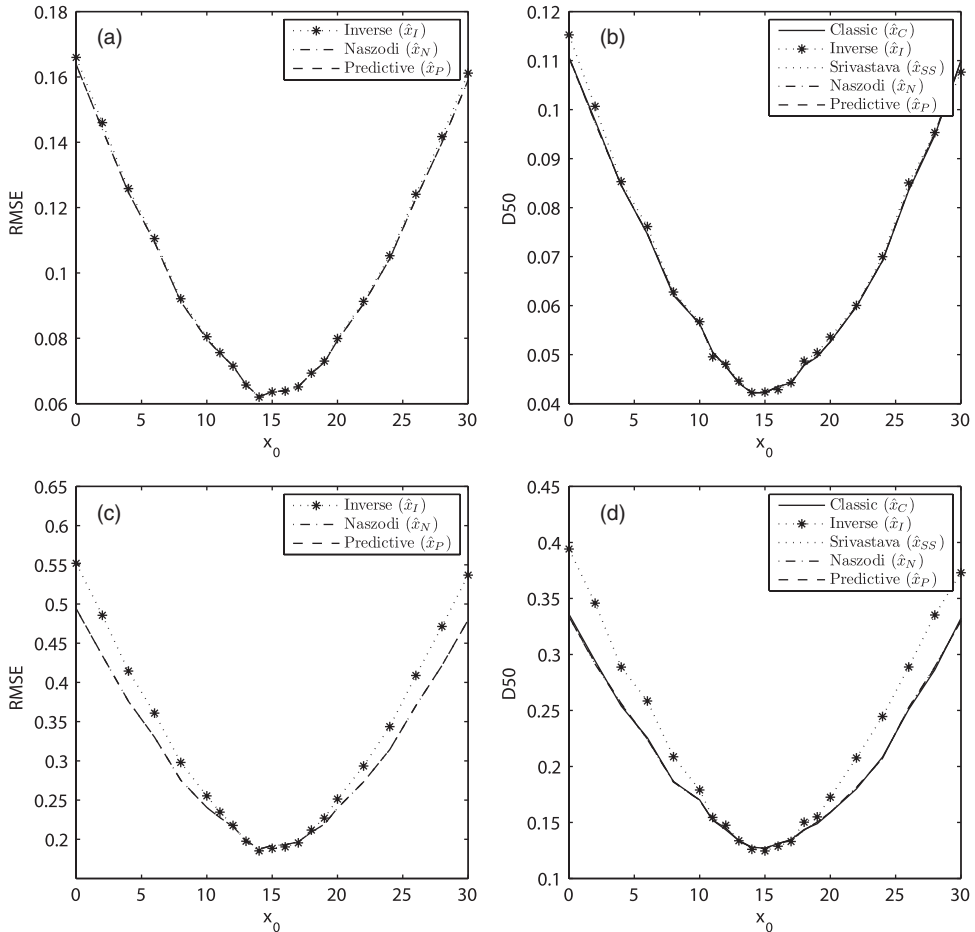


Figure 1. Curves of RMSE and $D50$ as functions of x_0 for the different estimators \hat{x}_I , \hat{x}_C , \hat{x}_{SS} , \hat{x}_N and \hat{x}_P in scenarios $M1$ ((a) and (b)) and $M2$ ((c) and (d)). Curves that are not visible are superposed to others such as curves corresponding to \hat{x}_N and \hat{x}_P in (c), and those corresponding to \hat{x}_C , \hat{x}_{SS} , \hat{x}_N and \hat{x}_P in (d).

estimators for extrapolation, while for interpolation, though they do not outperform \hat{x}_I , they provide improvement over \hat{x}_C and \hat{x}_{SS} . In such remarkably bad settings as $M4$, \hat{x}_N outperforms \hat{x}_P from the viewpoint of the RMSE.

However, the latter does not mean that \hat{x}_N performs as well as \hat{x}_P regarding other statistical criteria. In particular, from the viewpoint of the KL divergence from the true density $f(y; x_0, \gamma_0, \sigma_0^2)$ of a future observation Y , the plug-in density estimate $f(y; \hat{x}_N, \hat{\gamma}, \hat{\sigma}^2)$ provided by \hat{x}_N is worse than the predictive approximation $f_P(y; \hat{x}_P, \mathcal{D})$ based on \hat{x}_P , over the whole range of possible x_0 values. In this sense, \hat{x}_P yields a better estimate of the distribution of a future observation than \hat{x}_N . This is illustrated in Figure 3 where, for the setting $M4$, the expected KL divergences from $f(y; x_0, \gamma_0, \sigma_0^2)$ corresponding to the density estimates provided by \hat{x}_N and \hat{x}_P are plot as a function of x_0 . Note that the setting $M4$ is precisely the one in which \hat{x}_N showed the best performance relative to \hat{x}_P in terms of RMSE.

An attractive feature of the non-Bayesian predictive approach for calibration is that it provides not only point estimates of x_0 but also predictive likelihoods of all possible values of this parameter. This allows one to complement the point estimate \hat{x}_P with a likelihood-based appreciation of the

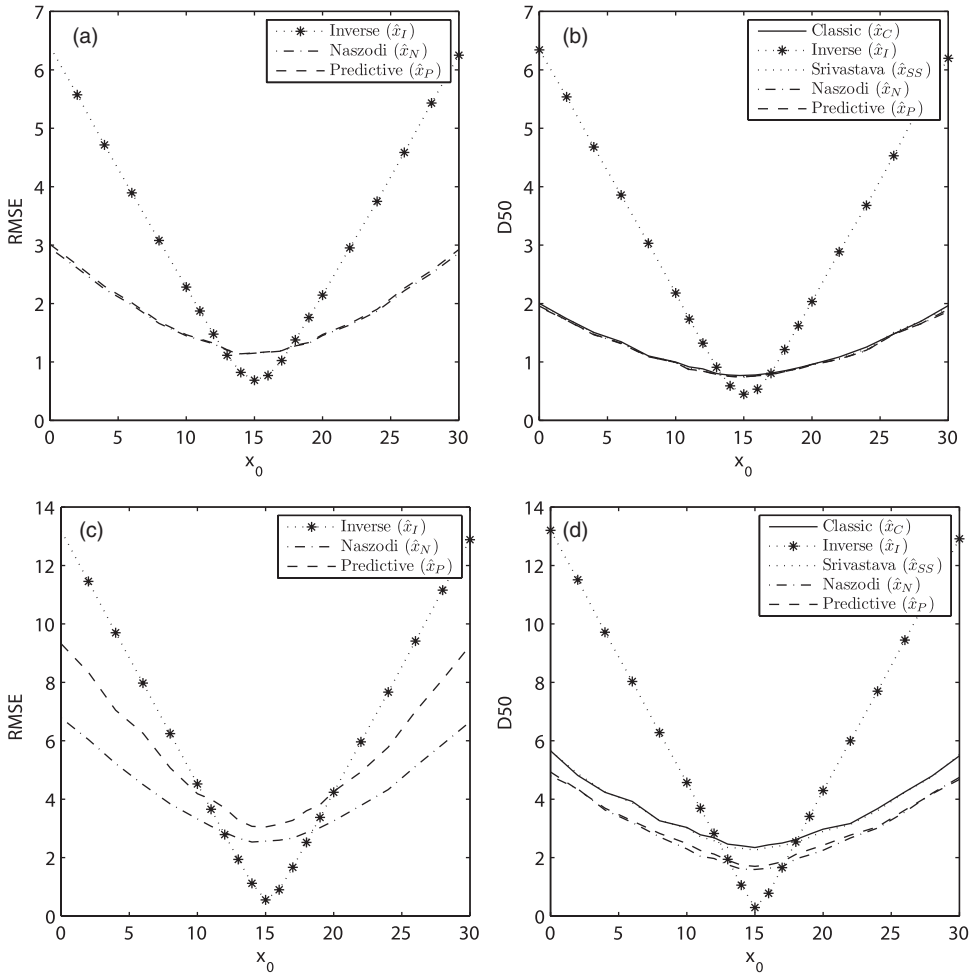


Figure 2. Curves of RMSE and $D50$ as functions of x_0 for the different estimators \hat{x}_I , \hat{x}_C , \hat{x}_{SS} , \hat{x}_N and \hat{x}_P in scenarios $M3$ ((a) and (b)) and $M4$ ((c) and (d)). Curves that are not visible are superposed to others such as curves corresponding to \hat{x}_C , \hat{x}_{SS} , \hat{x}_N and \hat{x}_P in (b), and those corresponding to \hat{x}_C and \hat{x}_{SS} in (d).

location of x_0 through the visual exploration of the relative predictive likelihood curve (for the importance of likelihood analysis in statistical inference, see e.g. [31,32]). As an illustration, Figure 4 presents the relative predictive likelihood curves for samples generated from setting $M1$ with $x_0 = 15$ (left-upper corner), and $x_0 = 6$ (right-upper corner); and samples from the setting $M3$ with $x_0 = 15$ (left-lower corner) and $x_0 = 6$ (right-lower corner). It can be observed that predictive likelihood curves show more precision in the assessment of x_0 (i.e., they are more spiky) for good settings than for bad settings (compare upper row vs. lower row in this Figure 4), as well as for x_0 values close to than far away from the centre of the calibration range (compare left vs. right columns in this figure).

To further illustrate this point, Figure 5 shows the predictive likelihood function for a new setting $M5$ in the extreme case in which $\beta = 0$. It is clear in this situation that the specific value of x_0 is indeterminate in the model, so it does not make sense to try to give any point estimate of x_0 . This is correctly reflected by the relative predictive likelihood function shown in Figure 5. Indeed, it assigns very high likelihoods to all possible values of x_0 .

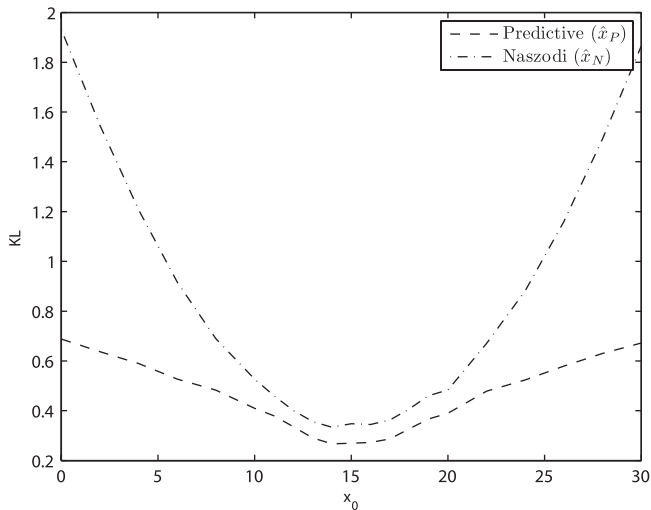


Figure 3. Curves of KL as a function of x_0 for the estimators \hat{x}_N and \hat{x}_P in scenario $M4$.

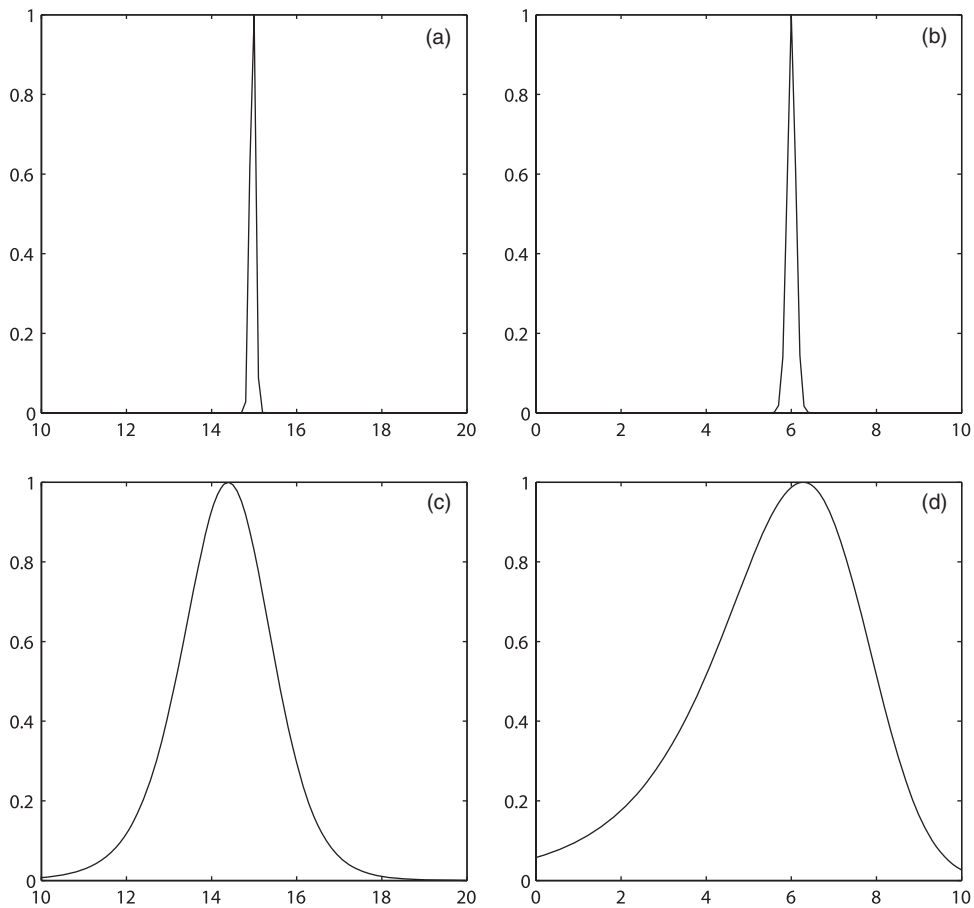


Figure 4. Relative predictive likelihood functions for different scenarios and x_0 values: (a) $M1$ with $x_0 = 15$, (b) $M1$ with $x_0 = 6$, (c) $M3$ with $x_0 = 15$ and (d) $M3$ with $x_0 = 6$.

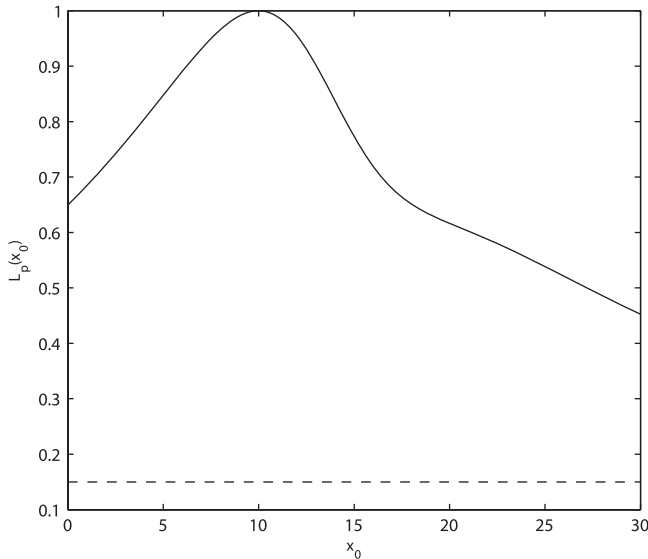


Figure 5. Relative predictive likelihood as a function of possible x_0 values in a setting in which $\beta = 0$. For reference, the cut-off 0.15 is indicated by the horizontal line, which corresponds to the 95% confidence interval according to the chi-squared approximation to the likelihood.

4. Concluding Remarks

In this paper, a new approach for statistical analysis of calibration problems has been introduced. This has been elaborated in detail in case of Gaussian univariate calibration problems to provide a non-Bayesian predictive likelihood function for the unknown calibration parameter x_0 and its point estimator \hat{x}_P .

The new estimator \hat{x}_P is consistent and has bounded MSE to zero-order asymptotic expansion (with respect to sample size n of the training data), likewise the classical estimator \hat{x}_C and the estimators \hat{x}_{SS} and \hat{x}_N proposed in [25,26], respectively.

Asymptotic analysis and simulation results with moderate sample sizes demonstrate that \hat{x}_P is better than \hat{x}_C and \hat{x}_{SS} over the whole range of possible x_0 values in the sense of both RMSE and $D50$ criteria.

For extrapolation problems (i.e., when x_0 is outside the calibration range), \hat{x}_P also outperforms the popular inverse estimator \hat{x}_I and has behaviour quite similar to \hat{x}_N . Thus, \hat{x}_P and \hat{x}_N are very appealing for extrapolation.

For interpolation problems (i.e., when x_0 is inside the calibration range), \hat{x}_P is worse than \hat{x}_I but not so much as \hat{x}_C and \hat{x}_{SS} .

The properties just summarized makes \hat{x}_P and \hat{x}_N appealing point estimators, especially for extrapolation.

In addition, the introduced approach have several advantageous features that are lacking in other Bayesian and non-Bayesian approaches to calibration. In particular:

- (a) It does not only provide point estimates but also makes it possible likelihood assessments by means of a non-Bayesian predictive likelihood function.
- (b) It can be extended to any (regular) parametric statistical model such as those which are multivariate, non-Gaussian or nonlinear.

These are two interesting lines of development of the non-Bayesian predictive approach to calibration that are worth of further research.

References

- [1] C. Osborne, *Statistical Calibration: A review*, Int. Statist. Rev. 59 (1991), pp. 309–336.
- [2] H. Martens and T. Naes, *Multivariate Calibration*, Wiley, Chichester, 1989.
- [3] P.J. Brown, *Measurement, Regression, and Calibration*, Clarendon Press, Oxford, 1993.
- [4] D.L. Massart, B.G.M. Vandeginste, L. Buydens, S.D. Jong, P.J. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier Science B. V., The Netherlands, 1997.
- [5] B.K. Lavine and J. Workman, *Fundamental reviews: chemometrics*, Anal. Chem. 74 (2002), pp. 2763–2770.
- [6] F.H. Walters and G.T. Rizzuto, *The calibration problem in statistics and its application to chemistry*, Anal. Lett. 21 (1988), pp. 2069–2076.
- [7] C. Eisenhart, *The interpretation of certain regression methods and their use in biological and industrial research*, Ann. Math. Statist. 10 (1939), pp. 162–186.
- [8] J. Berkson, *Estimation of a linear function for a calibration line; consideration of a recent proposal*, Technometrics 11 (1969), pp. 649–660.
- [9] E.J. Williams, *Regression methods in calibration problems*, Bull. Int. Statist. Inst. 43 (1969), pp. 17–28.
- [10] R.G. Krutchkoff, *Classical and inverse regression methods of calibration*, Technometrics 9 (1967), pp. 425–439.
- [11] R.G. Krutchkoff, *Classical and inverse regression methods of calibration in extrapolation*, Technometrics 11 (1969), pp. 605–608.
- [12] I.R. Dunsmore, *A Bayesian approach to calibration*, J. R. Statist. Soc. B. 30 (1968), pp. 396–405.
- [13] J. Aitchison and I.R. Dunsmore, *Statistical Prediction Analysis*, Cambridge University Press, London, 2002.
- [14] S.K. Perng, *A note on the inverse estimator for the linear calibration problem*, Commun. Statist. Theoret. Math 16 (1987), pp. 1743–1747.
- [15] T. Lwin and J.S. Maritz, *An analysis of the linear calibration controversy from the perspective of compound estimation*, Technometrics 24 (1982), pp. 235–242.
- [16] M. Halperin, *On inverse estimation in linear regression*, Technometrics 12 (1970), pp. 727–736.
- [17] C. Hagwood, *The calibration problem as an ill-posed inverse problem*, J. Statist. Plann. Inference 31 (1992), pp. 179–185.
- [18] G. Shukla, *On the problem of calibration*, Technometrics 14 (1972), pp. 547–553.
- [19] T.P. Turiel, G.J. Hahn, and W.T. Tucker, *New simulations results for the calibration and inverse median estimation problems*, Commun. Statist. Simul. C. 11 (1982), 677–713.
- [20] S. Chow and J. Shao, *On the difference between the classical and the inverse methods of calibration*, J. R. Statist. Soc. C Appl. 39 (1990), pp. 219–228.
- [21] R. Hüls, *Der mittlere quadratische Fehler verschiedener Kalibrationssch.*, Msc. diss., Dortmund University, 1996.
- [22] V. Centner, D.L. Massart, and S. Jongde, *Inverse calibration predicts better than classical calibration*, Fresenius J. Anal. Chem. 361 (1998), pp. 2–9.
- [23] J. Tellinghuisen, *Inverse vs. classical calibration for small data sets*, Fresenius J. Anal. Chem. 368 (2000), p. 585–588.
- [24] M.A. Ali and N. Singh, *An alternative estimator in inverse regression*, J. Statist. Comput. Simul. 14 (1981), pp. 1–15.
- [25] N. Singh and V.K. Srivastava, *Small disturbance asymptotic theory for linear calibration estimators*, Technometrics 31 (1989), pp. 373–378.
- [26] L.J. Naszödi, *Elimination of the bias in the course of calibration*, Technometrics 20 (1978), pp. 201–205.
- [27] I.R. Harris, *Predictive fit for natural exponential families*, Biometrika 76 (1989), pp. 675–684.
- [28] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
- [29] E. Williams, *A note on Regression methods in calibration*, Technometrics 11 (1969), pp. 189–192.
- [30] S. Kullback, *Information Theory and Statistics*, Dover Publications, New York, 1978.
- [31] D.A. Sprott, *Statistical inference in science*, Springer-Verlag, New York, 2000.
- [32] Y. Pawitan, *In All Likelihood. Statistical Modelling and Inference using Likelihood*, Clarendon Press, Oxford, 2001.

Appendix 1: Asymptotic expansion of the non-Bayesian predictive estimator

Denote $\hat{\mu}_0 = \mathbf{u}_0 \hat{\mathbf{y}}^\top = \hat{\alpha} + \hat{\beta} x_0$, $w_0 = \mathbf{u}_0 (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{u}_0^\top$ and $G = -\ln f_P(y_0; x_0, \mathcal{D})$. Taylor expansion of $f_{N(\hat{\mu}_0, w_0 \hat{\sigma}^2 + v/m)}$ (y_0) (defined by expression (9)) with respect to v around $\hat{\sigma}^2$ and taking into consideration that $(n-2)\hat{\sigma}^2/\sigma^2$ is a chi-square random variable with $n-2$ degrees of freedom it follows that

$$f_P(y_0; x_0, \mathcal{D}) = f_{N(\hat{\mu}_0, w_0 \hat{\sigma}^2 + \hat{\sigma}^2)}(y_0) + \frac{\partial^2}{\partial v^2} f_{N(\hat{\mu}_0, w_0 \hat{\sigma}^2 + v)}(y_0) \Big|_{v=\hat{\sigma}^2} \frac{\hat{\sigma}^4}{n-2} + R,$$

where the remainder term R is of order $1/n^2$. This and

$$w_0 = \frac{(x_0 - \bar{x})^2}{\sigma_x^2(n-1)} + \frac{1}{n}$$

imply that to first-order expansion with respect to $\varepsilon = 1/n$ it holds that

$$G \approx \frac{1}{2} \ln(2\pi) + \frac{(y_0 - \hat{\mu}_0)^2}{2\hat{\sigma}^2(\varepsilon((x_0 - \bar{x})^2/\sigma_x^2) + \varepsilon + 1)} + \frac{1}{2} \ln \left(\left(\frac{(x_0 - \bar{x})^2}{\sigma_x^2} \varepsilon + \varepsilon + 1 \right) \hat{\sigma}^2 \right) - \frac{3(y_0 - \hat{\mu}_0)^2}{2(\varepsilon((x_0 - \bar{x})^2/\sigma_x^2) + \varepsilon + 1)^3 \hat{\sigma}^2} \varepsilon + \frac{(y_0 - \hat{\mu}_0)^4}{4\hat{\sigma}^4(\varepsilon((x_0 - \bar{x})^2/\sigma_x^2) + \varepsilon + 1)^4} \varepsilon + \frac{3}{4(\varepsilon((x_0 - \bar{x})^2/\sigma_x^2) + \varepsilon + 1)^2} \varepsilon. \quad (\text{A1})$$

By definition, \hat{x}_P minimizes G . Thus,

$$\frac{\partial}{\partial x_0} G \Big|_{x_0 = \hat{x}_P} = 0 \quad (\text{A2})$$

for any $\varepsilon > 0$. At $\varepsilon = 0$ one has

$$G|_{\varepsilon=0} = -\ln f_{N(\hat{\mu}_0, \hat{\sigma}^2)}(y_0),$$

which is a function minimized by \hat{x}_C , so that

$$\hat{x}_P|_{\varepsilon=0} = \hat{x}_C.$$

Therefore, first-order Taylor expansion of \hat{x}_P with respect to ε gives

$$\hat{x}_P \approx \hat{x}_C + \frac{\partial}{\partial \varepsilon} \hat{x}_P \Big|_{\varepsilon=0} \varepsilon. \quad (\text{A3})$$

On the other hand, derivation of Equation (A2) with respect to ε leads to

$$\frac{\partial}{\partial \varepsilon} \hat{x}_P \Big|_{\varepsilon=0} = - \frac{(\partial/\partial \varepsilon)(\partial/\partial x_0)G|_{x_0 = \hat{x}_P}}{(\partial^2/\partial x_0^2)G|_{x_0 = \hat{x}_P}}. \quad (\text{A4})$$

From Equations (A1), (A3) and (A4) it follows that, to first-order expansion regarding ε , the non-Bayesian predictive estimator can be approximated by

$$\hat{x}_P \approx \bar{x} + \frac{(1 - \hat{\eta}_n^2)}{\hat{\beta}} (y_0 - \bar{y}), \quad (\text{A5})$$

where

$$\hat{\eta}_n^2 = \frac{\hat{\sigma}^2}{\sigma_x^2 \hat{\beta}^2 n}.$$

Appendix 2: Asymptotic expansion of the mean-square error of the non-Bayesian predictive estimator

Under the assumption of Gaussian errors in model (1), $(n-2)\hat{\sigma}^2/\sigma^2$ is a chi-square random variable with $n-2$ degrees of freedom, which is independent of the random vector $(\hat{\alpha}, \hat{\beta})$, and the latter has Gaussian distribution with mean (α, β) and covariance matrix $\sigma^2(\mathbf{U}^T \mathbf{U})^{-1}$. These facts and a first-order Taylor expansion of expression (A5) with respect to $(\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ around (α, β, σ) leads to the following approximation to the mean-square error of \hat{x}_P to terms of order $\varepsilon = 1/n$:

$$E(\hat{x}_P) \approx x_0, \quad (\text{A6})$$

$$V(\hat{x}_P) \approx \frac{\sigma^2}{\beta^2} + \eta_n^2 \left((x_0 - \bar{x})^2 - \frac{\sigma^2}{\beta^2} + \sigma_x^2 \right), \quad (\text{A7})$$

$$\text{MSE}(\hat{x}_P) = E(\hat{x}_P - x_0)^2 \approx \frac{\sigma^2}{\beta^2} + \eta^2 \left(\sigma_x^2 + (x_0 - \bar{x})^2 + 3 \frac{\sigma^2}{\beta^2} \right), \quad (\text{A8})$$

where

$$\eta_n^2 = \frac{\sigma^2}{\sigma_x^2 \beta^2 n}.$$

Appendix 3: Finiteness of the moments of the non-Bayesian predictive estimator

It will be shown here that the estimator \hat{x}_p studied through simulations in Section 3 has finite moments. For simplicity, without loss of generality, we will suppose that the intercept α in Equation (1) is zero. By definition, \hat{x}_p maximizes with respect to x the logarithm of the predictive density, i.e., it maximizes the function:

$$\ln g(x) = -\frac{1}{2} \frac{(y_0 - \hat{\beta}x)^2}{q\hat{\sigma}^2x^2 + \hat{\sigma}^2} - \frac{1}{2} \ln(q\hat{\sigma}^2x^2 + \hat{\sigma}^2),$$

where $q = 1/\sum_{i=1}^n x_i^2$.

Taking derivative of this function with respect to x and equating to zero it follows that \hat{x}_p is a root of the cubic polynomial equation:

$$a_0x^3 + a_1x^2 + a_2x + a_3 = 0, \quad (\text{A9})$$

where

$$\begin{aligned} a_0 &= q^2\hat{\sigma}^2, \\ a_1 &= y_0\hat{\beta}q, \\ a_2 &= \hat{\beta}^2 + q\hat{\sigma}^2 - y_0^2q, \\ a_3 &= -y_0\hat{\beta}. \end{aligned}$$

It is known that the roots of this equation have the form $x = (z + a_1)/3a_0$ where

$$z = \sqrt[3]{-\frac{N}{2} + \sqrt{\frac{N^2}{4} + \frac{M^3}{27}}} + \sqrt[3]{-\frac{N}{2} - \sqrt{\frac{N^2}{4} + \frac{M^3}{27}}},$$

and $M = 3(3a_0a_2 - a_1^2)$, $N = 2a_1^3 - 9a_0a_1a_2 + 27a_0^2a_3$.

Using the algebraic inequalities $\sqrt[3]{|u|} \leq (1 + |u|)$ and $\sqrt{|u|^2 + |v|^2} \leq |u| + |v|$ for any complex numbers u and v , where $|\cdot|$ denotes module of complex numbers, it can be shown that

$$|z| \leq 2 \left(1 + |N| + \frac{|M|^{3/2}}{\sqrt{27}} \right).$$

Recall that y_0 and $\hat{\beta}$ are normal random variables and $\hat{\sigma}^2$ is proportional to a chi-square distribution. Since M and N are polynomials in normal and chi-square random variables, $|z|$ have finite moments of all positive order. On the other hand, using the Cauchy–Schwartz inequality and the algebraic inequality $(|u| + |v|)^2 \leq 2(|u|^2 + |v|^2)$, it can be shown that

$$E|\hat{x}_p|^2 \leq \sqrt{\frac{1}{3} E \left(\frac{1}{a_0^4} \right)} \sqrt{4(E|z|^4 + Ea_1^4)}.$$

Since a_0 is proportional to a chi-square distribution, $E(1/a_0^4) < \infty$. This and the fact that $E|z|^4 < \infty$, $Ea_1^4 < \infty$ imply that,

$$E\hat{x}_p^2 < \infty.$$