Ann. Inst. Statist. Math.
Vol. 55, No. 1, 169–185 (2003)

PREDICTION AND CALIBRATION IN GENERALIZED LINEAR MODELS*

PAOLO VIDONI

Department of Statistics, University of Udine, via Treppo 18, I-33100 Udine, Italy,
e-mail: vidoni@dss.uniud.it

(Received May 1, 2001; revised December 3, 2001)

Abstract. This paper concerns prediction and calibration in generalized linear models. A new predictive procedure, giving improved prediction intervals, is briefly reviewed and further theoretical results, useful for calculations, are presented. Indeed, the calibration problem is faced within the classical approach and a suitable solution is obtained by inverting the associated improved prediction procedure. This calibration technique gives accurate confidence regions and it constitutes a substantial improvement over both the estimative solution and the naive solution, which involves, even for non-linear and non-normal models, the results available for the linear Gaussian case. Finally, some useful explicit formulae for the construction of prediction and calibration intervals are presented, with regard to generalized linear models with alternative error terms and link functions.

Key words and phrases: Asymptotic expansion, coverage probability, inverse prediction, prediction limit, predictive density.

1. Introduction

Prediction and calibration are usually viewed as related problems and, within the classical approach, calibration is considered as a kind of inverse prediction. In fact, calibration regions are usually obtained by inverting a suitable prediction procedure. This paper concerns prediction and calibration in generalized linear models. In particular, some recent results on prediction in generalized linear models are studied in more detail in order to obtain improved solutions for the prediction and the calibration problems.

Statistical calibration is, broadly speaking, an inferential procedure useful whenever two types of measurements or observations may be given for the same subject. Usually, the observations (x, ξ) and (y, z) refer to an accurate but expensive measure and to a cheap but less accurate measure, respectively, and the calibration procedure consists in the following two stages. In the first stage, called *calibration stage*, a training experiment, with both y and x available, is performed with the aim of defining a model for the relation between these two types of measurements. In the second stage, called *estimation (prediction) stage*, the observations z are used as indirect measurements of ξ, which is not available. That is, the model is considered in order to estimate the unknown values of ξ associated to further available observations z. Statistical calibration is extensively applied in Chemistry, Biology and Engineering. In this paper, we focus on *univariate* (y and x scalar) *controlled calibration*; namely, x is considered as fixed (non-random

*This research was partially supported by a grant from Ministero dell'Istruzione, dell'Università e della Ricerca, Italy.

and controlled) by design in the training experiment. Osborne (1991) provides an exhaustive review on statistical calibration; further general references are Brown (1993) and Sundberg (1999).

The classical approach to calibration goes back to Eisenhart (1939) and it aims to derive a suitable confidence interval for the unknown ξ. These confidence intervals are usually obtained by inverting prediction intervals, associated to the further available observation z. The coverage level of a confidence interval for ξ is equal to the coverage probability of the corresponding prediction interval for z. A serious problem, related to this approach, is that its use is mainly confined to normal linear models. Within different models, the confidence intervals for ξ are usually computed mimicking the normal linear case or using a simple estimative prediction procedure. However, the coverage level may be remarkably different from the target nominal value and then the calibration procedure is usually rather inaccurate. In this paper, the classical approach is extended in order to define a proper prediction-based calibration approach. This approach involves an improved prediction procedure and it gives a satisfactory solution to the above mention problems. There are different approaches to calibration, not considered here, based on likelihood methods (see Brown (1993), and Bellio (2000)) and on bootstrap techniques (Gruet *et al.* (1993)), which may provide good alternative solutions in a number of different models.

Within generalized linear models, Vidoni (2001) has recently defined a procedure, based on the notion of a predictive density, for the construction of prediction intervals with coverage probability equal to the nominal value to a close approximation. In this paper, we shall consider these recent results in order to obtain improved classical solutions for the calibration problem, useful as well for non-linear and non-normal models.

The paper is organized as follows. Section 2 gives a brief introduction to the prediction and the calibration problems and reviews the above mentioned results on prediction in generalized linear models. In Section 3, these results are complemented by giving a formal expression for the improved prediction limit and the associated distribution function and by considering the case with an unknown dispersion parameter. Finally, Section 4 presents some explicit formulae useful for calculation of improved prediction and calibration intervals for models with normal, gamma and Poisson distributed error terms.

2. Prediction and calibration in generalized linear models

2.1 *Preliminaries on prediction*

The prediction of the value of a future random variable, based on an observed sample, is usually expressed in terms of prediction intervals, or, equivalently, in terms of the predictive density which generates the required prediction intervals through its quantiles.

Let us assume that the observable random vector $Y = (Y_1, \ldots, Y_n)$ consists of random variables having marginal probability density functions $p_i(y;\omega)$, $i = 1, \ldots, n$, $\omega \in \Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}^+$, with respect to a suitable dominating measure. The future random variable $Z = Y_{n+1}$, independent of Y, has density $p_{n+1}(z;\omega)$, depending on the unknown parameter ω. Its observed value z is the interest quantity and prediction statements are usually based on $p_{n+1}(z;\omega)$ or, since ω in unknown, on a suitable estimator involving the observable sample Y. More precisely, we are interested in the definition of an α-prediction interval for z or, in particular, of an *α-prediction limit* $z_\alpha(y)$, such that,

exactly or approximately,

$$pr\{Z \leq z_\alpha(Y);\omega\} = \alpha, \tag{2.1}$$

for all $\omega \in \Omega$, where $\alpha \in (0,1)$ is fixed. The above probability is called *coverage probability* and refers to the joint distribution of (Y,Z). At least from the frequentist viewpoint, the goodness of an α-prediction limit is measured by the discrepancy between the associated coverage probability and the target value α.

In some particular cases, such as the normal linear regression model, it is possible to find prediction limits satisfying (2.1) exactly. This is usually done by considering suitable pivotal functions, that is functions of Y and Z, whose distribution is free of ω. In general, this is not possible and we define prediction limits which constitute approximate solutions to (2.1). Since ω is unknown, the simplest procedure is that one giving the *estimative prediction limit* $\hat{z}^e_\alpha = z^e_\alpha(\hat{\omega})$, obtained by substituting ω with an asymptotically efficient estimator $\hat{\omega} = \hat{\omega}(Y)$ in $z_\alpha(\omega)$. Hereafter, $\hat{\omega}$ is the maximum likelihood estimator. Furthermore, $z_\alpha(\omega)$ is the α-quantile associated with $p_{n+1}(z;\omega)$, while $\hat{z}^e_\alpha$ is the α-quantile associated with the *estimative predictive density* $p_{n+1}(z;\hat{\omega})$.

For continuous random variables, Barndorff-Nielsen and Cox (1996) emphasize that the coverage probability associated to $\hat{z}^e_\alpha$ is $\alpha+O(n^{-1})$ and define a procedure giving prediction limits which satisfy (2.1), in the conditional form, to order $O(n^{-3/2})$. However, these limits are implicitly defined and the associated predictive density is usually not in a closed-form. Vidoni (1998) introduces a prediction limit, expressed as a modification of the estimative one, which is asymptotically equivalent, up to terms of order $O(n^{-3/2})$, to that of Barndorff-Nielsen and Cox (1996) and hence meets the same asymptotic properties. Moreover, the associated predictive density has a relatively simple closed-form expression, which may be useful for computations. Vidoni (2001) shows that this predictive density gives improved prediction limits for discrete random variables as well.

2.2 *Improved prediction in generalized linear models*

Let us suppose that $Y_1,\ldots,Y_n,Y_{n+1}$, with $Z = Y_{n+1}$, are mutually independent and such that Y_i, $i = 1,\ldots,n+1$ has probability density function, with respect to a suitable dominating measure,

$$p_i(y_i;\omega,\lambda) = c(\lambda,y_i)\exp(\lambda\{y_i\theta_i - K(\theta_i)\}), \qquad y_i \in \mathcal{Y} \subseteq I\!R,$$

where the parameter θ_i is a linear function $\theta_i = x_i^T\omega$, called *linear predictor*, of a d-dimensional vector $\omega = (\omega_1,\ldots,\omega_d)^T$ of unknown parameters and a d-dimensional vector $x_i = (x_{i,1},\ldots,x_{i,d})^T$ of known covariates. The parameter $\lambda \in \Lambda \subseteq I\!R^+$ is the *index parameter*, whereas $\sigma^2 = 1/\lambda$ is the *dispersion parameter*. Thus, Y_i follows a reproductive exponential dispersion model (see Jørgensen (1997), Chapter 3) with mean $\mu_i = \mu(\theta_i) = dK(\theta_i)/d\theta_i$ and variance $\sigma^2V(\mu_i)$, where the *variance function* $V(\mu_i) = d^2K(\theta_i)/d\theta_i^2|_{\theta_i=\theta(\mu_i)}$, with $\theta(\cdot)$ the inverse of $\mu(\cdot)$. A generalization of this model, giving the class of generalized linear models (see McCullagh and Nelder (1989)), is obtained by considering a monotonic differentiable *link function* $g(\cdot)$ such that $g(\mu_i) = x_i^T\omega$. Moreover, $\mu_i = g^{-1}(x_i^T\omega)$ and $\theta_i = \theta(g^{-1}(x_i^T\omega))$, with $g^{-1}(\cdot)$ the inverse of $g(\cdot)$. The simplified model with $\theta_i = x_i^T\omega$ is obtained whenever the *canonical* link function $g(\cdot) = \theta(\cdot)$ is considered.

Let us consider σ^2 known or estimated. Since the parameters σ^2 and ω are orthogonal, this assumption does not influence the maximum likelihood estimator $\hat{\omega}$. The

extension to the case with both ω and σ^2 unknown requires an additional computational effort and it will be presented in Subsection 3.2. Hereafter, we use index notation and Einstein summation convention, according to which if an index occurs more than once in a summand then summation over that index is understood. The convention is suppressed only for the index $i = 1, \ldots, n+1$, which labels the observations.

The predictive density, which gives improved prediction limits both in the continuous and in the discrete case, may be specified as follows (see Vidoni (2001), for a complete derivation)

$$p^{+}_{n+1}(z; \tilde{\omega}_\alpha, \lambda) = p_{n+1}(z; \tilde{\omega}_\alpha, \lambda)\{1 + H_{n+1}(z; \tilde{\omega}_\alpha, \lambda)\}, \tag{2.2}$$

where $H_{n+1}(z; \tilde{\omega}_\alpha, \lambda)$ is

$$H_{n+1}(z; \omega, \lambda) = -\frac{1}{2}\lambda[\lambda(z - \mu_{n+1})^2 \theta_{n+1,r}\theta_{n+1,s} - (z - \mu_{n+1}) \{\theta_{n+1,r}(\upsilon_{s,tu} + \upsilon_{s,t,u})i^{tu} - \theta_{n+1,rs}\} - \theta_{n+1,r}\mu_{n+1,s}]i^{rs},$$

computed at $\omega = \tilde{\omega}_\alpha$ and

$$\tilde{\omega}_{\alpha r} = \hat{\omega}_r + \frac{\lambda(\hat{z}^e_\alpha - \hat{\mu}_{n+1})x_{n+1,s}\hat{i}^{rs}}{V(\hat{\mu}_{n+1})g'(\hat{\mu}_{n+1})}, \qquad r = 1, \ldots, d.$$

Here, $i^{rs} = i^{rs}(\omega)$, $r, s = 1, \ldots, d$, is the (r, s)-element of the inverse of the expected information matrix, $\upsilon_{s,tu} = E(\ell_s \ell_{tu})$ and $\upsilon_{s,t,u} = E(\ell_s \ell_t \ell_u)$, $s, t, u = 1, \ldots, d$, where $\ell_s = \partial \ell(\omega; y)/\partial \omega_s$ and $\ell_{tu} = \partial^2 \ell(\omega; y)/\partial \omega_t \partial \omega_u$, with $\ell(\omega; y)$ the loglikelihood function. In this case,

$$\upsilon_{s,tu} = \lambda \sum_{i=1}^{n} \frac{x_{i,s}\theta_{i,tu}}{g'(\mu_i)}, \quad \upsilon_{s,t,u} = \lambda \sum_{i=1}^{n} K'''(\theta(\mu_i))\theta_{i,s}\theta_{i,t}\theta_{i,u}, \quad i_{rs} = \lambda \sum_{i=1}^{n} \frac{x_{i,r}x_{i,s}}{\{g'(\mu_i)\}^2 V(\mu_i)}.$$

Furthermore, $K'''(\theta(\mu_i)) = d^3K(\theta_i)/d\theta_i^3|_{\theta_i = \theta(\mu_i)}$ and

$$\mu_{i,s} = \partial \mu_i/\partial \omega_s = x_{i,s}\{g'(\mu_i)\}^{-1}, \qquad \theta_{i,r} = \partial \theta_i/\partial \omega_r = x_{i,r}\{V(\mu_i)g'(\mu_i)\}^{-1},$$
$$\theta_{i,rs} = \partial^2 \theta_i/\partial \omega_r \partial \omega_s = -x_{i,r}x_{i,s}[K'''(\theta(\mu_i))\{V(\mu_i)\}^{-3}\{g'(\mu_i)\}^{-2} + g''(\mu_i)\{V(\mu_i)\}^{-1}\{g'(\mu_i)\}^{-3}],$$

with $i = 1, \ldots, n+1$, where $g'(\cdot)$ and $g''(\cdot)$ are the first and the second derivatives of $g(\cdot)$. With the canonical link function, we have that $\upsilon_{s,tu} = 0$, $i_{rs} = \lambda \sum_{i=1}^{n} V(\mu_i)x_{i,r}x_{i,s}$, $\upsilon_{s,t,u} = \lambda \sum_{i=1}^{n} K'''(\theta(\mu_i))x_{i,s}x_{i,t}x_{i,t}$, $\mu_{i,s} = x_{i,s}V(\mu_i)$, $\theta_{i,r} = x_{i,r}$ and $\theta_{i,rs} = 0$. The above quantities are evaluated at $\mu_i = g^{-1}(x_i^T\omega)$, $i = 1, \ldots, n+1$. Note that, whenever the order of the expected likelihood quantities i_{rs}, $\upsilon_{s,tu}$ and $\upsilon_{s,t,u}$ is $O(n)$, then $H_{n+1}(z; \tilde{\omega}_\alpha, \lambda)$ is $O(n^{-1})$ and $\tilde{\omega}_{\alpha r} = \hat{\omega}_r + O(n^{-1})$, $r = 1, \ldots, d$. In (2.2), $H_{n+1}(z; \tilde{\omega}_\alpha, \lambda)$ may be substituted by $H_{n+1}(z; \hat{\omega}, \lambda)$ without a change in the order of approximation. Here and in what follows, a hat indicates evaluation at $\omega = \hat{\omega}$, and a tilde evaluation at $\omega = \tilde{\omega}_\alpha$.

Function (2.2), which is a modification of the estimative density $p_{n+1}(z; \tilde{\omega}_\alpha, \lambda)$, depends, through the estimator $\tilde{\omega}_\alpha$, on the specified α. The predictive procedure proposed by Barndorff-Nielsen and Cox (1996) does not present this drawback. However, this is the price to be paid in order to obtain a simplified predictive density. Since the aim here is to derive suitable prediction limits for a fixed value α, the predictive density (2.2) turns out to be useful as well, especially for computations.

2.3 *Preliminaries on calibration*

Let us consider the controlled calibration problem for generalized linear models. More precisely, we adopt for $Y_1, \ldots, Y_n, Z$, with $Z = Y_{n+1}$, the general assumptions of Subsection 2.2. Furthermore, we suppose that the calibration stage consists of a training experiment giving the observations $y = (y_1, \ldots, y_n)$, with the corresponding covariates x_i, $i = 1, \ldots, n$, known and fixed by design. In the prediction stage, an additional observation $z = y_{n+1}$ is available, while the associated covariate $\xi = x_{n+1}$ is unknown. Thus, x_{n+1} is here the interest quantity and the aim is to derive suitable estimators or confidence regions for x_{n+1}, based on the observations (y, z). We focus on the univariate case, by assuming $\omega = (\omega_1, \omega_2)^T$ and $x_i = (1, x_{i,2})^T$, $i = 1, \ldots, n+1$; for ease of exposition, sometimes we consider the simpler case with scalar ω and x_i, $i = 1, \ldots, n+1$.

Within the classical approach, confidence intervals for x_{n+1} are obtained by inverting prediction intervals for Z. More precisely, if $(z_1(\hat{\omega}, \sigma; x_{n+1}), z_2(\hat{\omega}, \sigma; x_{n+1}))$ is an α-prediction interval for z, a confidence region for the unknown covariate x_{n+1} is given by $\{x_{n+1} \in I\!R : z_1(\hat{\omega}, \sigma; x_{n+1}) \leq z \leq z_2(\hat{\omega}, \sigma; x_{n+1})\}$, where z is now available. The associated coverage level is equal to the coverage probability of the corresponding prediction interval. Sometimes, the inversion procedure may not result in an interval; moreover, the confidence region may be unsatisfactory or even meaningless, but this usually happens when the model is weakly confirmed by the data.

Let us start by reviewing the well known results on calibration for simple linear regression models. We assume that Y_i, $i = 1, \ldots, n+1$, follows a $N(\omega_1 + \omega_2 x_{i,2}, \sigma^2)$ distribution, with σ^2 known. The random variable $\{Z - (\hat{\omega}_1 + \hat{\omega}_2 x_{n+1,2})\}/\{\sigma c(x_{n+1,2})\}$ is a pivotal quantity with a $N(0,1)$ distribution. Here, $\hat{\omega}_1$ and $\hat{\omega}_2$ are the maximum likelihood estimators for ω_1 and ω_2 and $c(x_{n+1,2}) = \{1 + 1/n + (x_{n+1,2} - \bar{x})^2/S_{xx}\}^{1/2}$, with $\bar{x}$ and S_{xx} the sample mean and the sample deviance of $x_{i,2}$, $i = 1, \ldots, n$. Whenever σ is unknown, it can be substituted by $\hat{\sigma}\sqrt{n/(n-2)}$, with $\hat{\sigma}$ the maximum likelihood estimator, giving a t-distribution with $n-2$ degrees of freedom. An exact α-prediction interval for Z can be easily defined by means of the above pivotal quantity. In order to derive a confidence interval for x_{n+1}, with an exact confidence level α, we need to invert this prediction interval, by considering the actual available further observation z. The inversion procedure usually gives a suitable confidence interval for x_{n+1}. It is well-known that this procedure may result in unsatisfactory confidence regions, but this is mainly confined to the case when the slope parameter ω_2 is not statistically different from zero, namely, when $|\hat{\omega}_2\sqrt{S_{xx}}/\sigma| < u_{(1+\alpha)/2}$, with $u_{(1+\alpha)/2}$ the $(1+\alpha)/2$-quantile of the standard normal distribution, or, in general, when the model does not provide a good description of the data available (see Brown (1993), § 2.3).

This approach holds for the linear Gaussian case. In non-linear and non-normal models, confidence intervals for x_{n+1} are usually derived by mimicking the results available for the normal linear case. For example, in generalized linear models with normal error and link function $g(\mu)$, different from the canonical one, we can consider an approximate pivotal quantity $\{Z - g^{-1}(x_i^T\hat{\omega})\}/(\sigma^2 + \hat{V}_g)^{1/2}$, with $\hat{V}_g$ a suitable estimator for the variance of $g^{-1}(x_i^T\hat{\omega})$. As a general alternative, an estimative predictive procedure can be taken into account. That is, an α-confidence interval for x_{n+1} is derived by inverting $(z^e_{(1-\alpha)/2}(\hat{\omega}, \sigma; x_{n+1}), z^e_{(1+\alpha)/2}(\hat{\omega}, \sigma; x_{n+1}))$, namely the estimative α-prediction interval for Z based on the the maximum likelihood estimator $\hat{\omega}$. However, the estimative solution is usually rather inaccurate, since the coverage probability of the prediction

interval, and therefore the actual confidence level of the associated confidence interval, may be remarkably different from the nominal value α. Indeed, the solution based on an approximate pivotal quantity may present unsatisfactory results as well, especially for highly non-linear regression models.

In the following, we shall consider an improved classical solution for the calibration problem. More precisely, this procedure consists in inverting an α-prediction interval based on the predictive density (2.2), which presents coverage probability closed to the target value α. Thus, the associated confidence region for x_{n+1}, which has an actual confidence level closed to the nominal value α, is given by $\{x_{n+1} \in I\!R : z^+_{(1-\alpha)/2}(\tilde{\omega}_{(1-\alpha)/2}, \sigma; x_{n+1}) \leq z \leq z^+_{(1+\alpha)/2}(\tilde{\omega}_{(1+\alpha)/2}, \sigma; x_{n+1})\}$, where $z^+_\gamma(\tilde{\omega}_\gamma, \sigma; x_{n+1})$ is the γ-quantile of a distribution with density (2.2). When an explicit expression for the confidence region does not exists, a solution may be obtained by means of a suitable numeric inversion procedure. Moreover, as in the linear normal case, the confidence region may be unsatisfactory or even meaningless, but this usually occurs when the model is weekly confirmed by the data or concerns an asymptotically negligible subset of the region, outside the design of the training experiment.

A final point regards calibration within a discrete model. Vidoni (2001) emphasizes that, in this case, the coverage probability of the estimative prediction interval consists of an unavoidable part due to discreteness and a further part related to the estimative procedure. Indeed, prediction intervals based on (2.2) improve the estimative ones, since, neglecting the error term due to discreteness, their coverage probability equals the target value to third order accuracy. Therefore, the same improvement maintains for the actual level of the associated confidence region for x_{n+1}. In this framework, besides interval estimation, it may be of interest to estimate the value (or the set of values) for x_{n+1}, such that a fixed potential observation $z \in \mathcal{Y}$ corresponds to the α-quantile of the future random variable Z. For example, given a suitable observation $z \in \mathcal{Y}$, the set $\{x_{n+1} \in I\!R : z^+_{1/2}(\tilde{\omega}_{1/2}, \sigma; x_{n+1}) = z\}$ defines the covariate values for which z may be reasonably considered as the median observation in the prediction stage. With regard to bioassay applications, these estimates for x_{n+1} can be viewed as a general alternative to the notion of α-level effective or lethal dose.

3. Further theoretical results on improved prediction

3.1 *Useful formulae for prediction intervals and distributions*

In this section we consider the problem of calculation of prediction intervals in generalized linear models. In particular, we focus on the derivation of explicit expressions for the improved α-prediction limit and the associated distribution function.

Let us start by noticing that the modifying term $H_{n+1}(z; \omega, \lambda)$ can be rewritten as a linear combination of the first two orthogonal polynomials (see Barndorff-Nielsen and Cox (1989), § 1.6) associated to $p_{n+1}(z; \omega, \lambda)$, with coefficient express in terms of cumulants, given by

$$B_1(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})}{\sigma\{V(\mu_{n+1})\}^{1/2}}, \tag{3.1}$$

$$B_2(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})^2}{\sigma^2 V(\mu_{n+1})} - \frac{\sigma K'''(\theta(\mu_{n+1}))}{\{V(\mu_{n+1})\}^{3/2}} \frac{(z - \mu_{n+1})}{\sigma\{V(\mu_{n+1})\}^{1/2}} - 1. \tag{3.2}$$

More precisely, we have that

$$H_{n+1}(z;\omega,\lambda) = H(z,x_{n+1};\omega,\sigma) = \frac{1}{2}A_1(x_{n+1};\omega,\sigma)B_1(z,x_{n+1};\omega,\sigma) - \frac{1}{2}A_2(x_{n+1};\omega,\sigma)B_2(z,x_{n+1};\omega,\sigma), \tag{3.3}$$

where

$$A_1(x_{n+1};\omega,\sigma) = \frac{x_{n+1,r}x_{n+1,s}g''(\mu_{n+1})i^{rs}}{\sigma\{V(\mu_{n+1})\}^{1/2}\{g'(\mu_{n+1})\}^3} + \frac{x_{n+1,r}(v_{s,tu}+v_{s,t,u})i^{tu}i^{rs}}{\sigma\{V(\mu_{n+1})\}^{1/2}g'(\mu_{n+1})}, \tag{3.4}$$

$$A_2(x_{n+1};\omega,\sigma) = \frac{x_{n+1,r}x_{n+1,s}i^{rs}}{\sigma^2 V(\mu_{n+1})\{g'(\mu_{n+1})\}^2}, \tag{3.5}$$

with the obvious simplification whenever the canonical link function is considered.

The α-prediction limit, based on the modified predictive density (2.2), is defined as the quantity $\tilde{z}_\alpha^+ = z_\alpha^+(\tilde{\omega}_\alpha,\sigma;x_{n+1})$ such that, in the continuous case, $F_{n+1}^+(\tilde{z}_\alpha^+;\tilde{\omega}_\alpha,\lambda)=\alpha$ or, in the discrete case, $\tilde{z}_\alpha^+ = \inf\{z\in\mathcal{Y} : F_{n+1}^+(z;\tilde{\omega}_\alpha,\lambda)\geq\alpha\}$. Here, $\mathcal{Y}$ is the support of Z and $F_{n+1}^+(z;\tilde{\omega}_\alpha,\lambda)$ is the distribution function associated to (2.2). If the modifying term is considered in the alternative form (3.3), we have that, in the continuous case,

$$F_{n+1}^+(z;\tilde{\omega}_\alpha,\lambda) = F_{n+1}(z;\tilde{\omega}_\alpha,\lambda) + \frac{1}{2}A_1(x_{n+1};\tilde{\omega}_\alpha,\sigma)\int_{-\infty}^{z} B_1(u,x_{n+1};\tilde{\omega}_\alpha,\sigma)p_{n+1}(u;\tilde{\omega}_\alpha,\lambda)du - \frac{1}{2}A_2(x_{n+1};\tilde{\omega}_\alpha,\sigma)\int_{-\infty}^{z} B_2(u,x_{n+1};\tilde{\omega}_\alpha,\sigma)p_{n+1}(u;\tilde{\omega}_\alpha,\lambda)du, \tag{3.6}$$

where $F_{n+1}(z;\tilde{\omega}_\alpha,\lambda)$ is the estimative distribution function based on $\tilde{\omega}_\alpha$ and $B_1(\cdot)$, $B_2(\cdot)$, $A_1(\cdot)$ and $A_2(\cdot)$, are (3.1), (3.2), (3.4) and (3.5) computed at $\omega=\tilde{\omega}_\alpha$. The discrete case is recovered by substituting the integration sign with the summation sign.

From (3.6), it is easy to show that, to the relevant order of approximation, the improved α-prediction limit $\tilde{z}_\alpha^+$, for continuous random variables, may be expressed as a suitable modification of the estimative prediction limit based on $\tilde{\omega}_\alpha$, namely $\tilde{z}_\alpha^e = z_\alpha^e(\tilde{\omega}_\alpha,\sigma;x_{n+1})$. That is, to third order accuracy,

$$\tilde{z}_\alpha^+ = \tilde{z}_\alpha^e - \frac{1}{2}\{p_{n+1}(\tilde{z}_\alpha^e;\tilde{\omega}_\alpha,\lambda)\}^{-1}A_1(x_{n+1};\tilde{\omega}_\alpha,\sigma)\cdot\int_{-\infty}^{\tilde{z}_\alpha^e} B_1(z,x_{n+1};\tilde{\omega}_\alpha,\sigma)p_{n+1}(z;\tilde{\omega}_\alpha,\lambda)dz + \frac{1}{2}\{p_{n+1}(\tilde{z}_\alpha^e;\tilde{\omega}_\alpha,\lambda)\}^{-1}A_2(x_{n+1};\tilde{\omega}_\alpha,\sigma)\cdot\int_{-\infty}^{\tilde{z}_\alpha^e} B_2(z,x_{n+1};\tilde{\omega}_\alpha,\sigma)p_{n+1}(z;\tilde{\omega}_\alpha,\lambda)dz. \tag{3.7}$$

Expression (3.7) holds for continuous random variables. Maintaining the same order for the error, $F_{n+1}^+(z;\tilde{\omega}_\alpha,\lambda)$ and $\tilde{z}_\alpha^+$ can be further approximated by suitable modifications of the estimative distribution function and the estimative prediction limit based on $\hat{\omega}$. We obtain, respectively, the approximations

$$
\begin{aligned}
(3.8)\quad F^{\circ}_{n+1}(z;\hat{\omega},\lambda,\alpha) &= F_{n+1}(z;\hat{\omega},\lambda) \\
&\quad - \frac{1}{2}A_2(x_{n+1};\hat{\omega},\sigma)\int_{-\infty}^{z} B_2(u,x_{n+1};\hat{\omega},\sigma)p_{n+1}(u;\hat{\omega},\lambda)du \\
&\quad + \left\{\frac{1}{2}A_1(x_{n+1};\hat{\omega},\sigma) + B_1(\hat{z}^e_\alpha,x_{n+1};\hat{\omega},\sigma)A_2(x_{n+1};\hat{\omega},\sigma)\right\} \\
&\quad \cdot \int_{-\infty}^{z} B_1(u,x_{n+1};\hat{\omega},\sigma)p_{n+1}(u;\hat{\omega},\lambda)du
\end{aligned}
$$

and

$$
\begin{aligned}
(3.9)\quad \hat{z}^{\circ}_\alpha = z^{\circ}_\alpha(\hat{\omega},\sigma;x_{n+1}) &= \hat{z}^e_\alpha + \hat{z}^e_{\alpha/r}B_1(\hat{z}^e_\alpha,x_{n+1};\hat{\omega},\sigma)\frac{x_{n+1,s}\hat{\imath}^{rs}}{\sigma\{V(\hat{\mu}_{n+1})\}^{1/2}g'(\hat{\mu}_{n+1})} \\
&\quad - \frac{1}{2}\{p_{n+1}(\hat{z}^e_\alpha;\hat{\omega},\lambda)\}^{-1}A_1(x_{n+1};\hat{\omega},\sigma)\int_{-\infty}^{\hat{z}^e_\alpha} B_1(z,x_{n+1};\hat{\omega},\sigma)p_{n+1}(z;\hat{\omega},\lambda)dz \\
&\quad + \frac{1}{2}\{p_{n+1}(\hat{z}^e_\alpha;\hat{\omega},\lambda)\}^{-1}A_2(x_{n+1};\hat{\omega},\sigma)\int_{-\infty}^{\hat{z}^e_\alpha} B_2(z,x_{n+1};\hat{\omega},\sigma)p_{n+1}(z;\hat{\omega},\lambda)dz.
\end{aligned}
$$

Here, $F_{n+1}(z;\hat{\omega},\lambda)$ and $\hat{z}^e_\alpha = z^e_\alpha(\hat{\omega},\sigma;x_{n+1})$ are the estimative distribution function and the estimative prediction limit based on the maximum likelihood estimator $\hat{\omega}$ and $\hat{z}^e_{\alpha/r} = \partial\hat{z}^e_\alpha/\partial\hat{\omega}_r$, $r = 1,\ldots,d$. Indeed, $B_1(\cdot)$, $B_2(\cdot)$, $A_1(\cdot)$, $A_2(\cdot)$, are (3.1), (3.2), (3.4), (3.5) computed at $\omega = \hat{\omega}$. As before, the discrete case is recovered in (3.8) by substituting the integration sign with the summation sign and then, to the relevant order of approximation, $\tilde{z}^+_\alpha$ equals $\hat{z}^{\circ}_\alpha = \inf\{z \in \mathcal{Y} : F^{\circ}_{n+1}(z;\hat{\omega},\lambda,\alpha) \geq \alpha\}$. Expression (3.9) holds for continuous random variables. Formula (3.8) is obtained by means of a Taylor expansion for $F_{n+1}(z;\tilde{\omega}_\alpha,\lambda)$ around $\tilde{\omega}_\alpha = \hat{\omega}$ and it implicitly requires that the order of integration and differentiation can be interchanged in $\partial F_{n+1}(z;\hat{\omega},\lambda)/\partial\hat{\omega}_r$, $r = 1,\ldots,d$. Formula (3.9) is a consequence of a general result given by Vidoni ((1998), § 3).

Relations (3.6)–(3.9) are useful for the computation of improved prediction intervals and distribution functions. However, since the above formulae involve integration or summation, a final closed-form expression is usually not available. In Section 4, we shall consider some generalized linear models where simple explicit results may be obtained. Finally, a further important point concerns the order of the error term in the improved predictive density (2.2) and in the expressions for prediction limits and distribution functions given in this section. If the order of the expected likelihood quantities i_{rs}, $\upsilon_{s,tu}$, $\upsilon_{s,t,u}$ is not $O(n)$, it is not necessarily true that the prediction limits based on (2.2) present coverage probability equal to the target value to third order accuracy. Moreover, the error term, which usually depends on the future covariate x_{n+1}, may not be negligible, especially when the position of x_{n+1} is far from the center of the x_i, $i = 1,\ldots,d$, considered in the training experiment. Thus, a careful investigation of the size of the error term is useful at least in more problematic situations.

3.2 *The case with unknown dispersion parameter*

In this section we complement the results on prediction in generalized linear models, by computing the modified predictive density $p^+_{n+1}(\cdot)$ in the general case where the dispersion parameter σ^2 is unknown. Following Vidoni (1998), the predictive density, which gives improved prediction limits both in the continuous and in the discrete case,

is

$$p^{+}_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha) = p_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)\{1 + H_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha) + S_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)\}, \tag{3.10}$$

where $S_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$ is

$$S_{n+1}(z;\omega,\lambda) = -\frac{1}{2}[(\ell_{n+1;\sigma^2})^2 - \ell_{n+1;\sigma^2}\{(\upsilon_{\sigma^2,tu} + \upsilon_{\sigma^2,t,u})i^{tu} + (\upsilon_{\sigma^2,\sigma^2\sigma^2} + \upsilon_{\sigma^2,\sigma^2,\sigma^2})i^{-1}_{\sigma^2\sigma^2}\} + \ell_{n+1;\sigma^2\sigma^2}]i^{-1}_{\sigma^2\sigma^2},$$

computed at $(\omega,\lambda) = (\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$, with $\tilde{\lambda}_\alpha = 1/\tilde{\sigma}^2_\alpha$, $\tilde{\sigma}^2_\alpha = \hat{\sigma}^2 + \ell_{n+1;\sigma^2}(\hat{\omega},1/\hat{\sigma}^2;\hat{z}^e_\alpha)i^{-1}_{\sigma^2\sigma^2}$ and $\hat{\sigma}^2$ the maximum likelihood estimator for σ^2. Here, $H_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$ is the modifying term given is Subsection 2.2 computed at $(\omega,\lambda) = (\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$, with $\tilde{\omega}_\alpha$ the well-known modification of the maximum likelihood estimator for ω. In this section, the index notation is considered only for the components of the d-dimensional parameter ω, which is orthogonal to σ^2. Moreover,

$$\ell_{i;\sigma^2} = \ell_{i;\sigma^2}(\omega,1/\sigma^2;y_i) = \frac{\partial \log p_i(y_i;\omega,1/\sigma^2)}{\partial\sigma^2} = -\frac{y_i\theta_i - K(\theta_i)}{\sigma^4} - \frac{c'(1/\sigma^2,y_i)}{\sigma^4 c(1/\sigma^2,y_i)},$$

$$\ell_{i;\sigma^2\sigma^2} = \frac{\partial^2 \log p_i(y_i;\omega,1/\sigma^2)}{\partial\sigma^2\partial\sigma^2} = -2\frac{\ell_{i;\sigma^2}}{\sigma^2} + \frac{1}{\sigma^8}\left[\frac{c''(1/\sigma^2,y_i)}{c'(1/\sigma^2,y_i)} - \left\{\frac{c'(1/\sigma^2,y_i)}{c(1/\sigma^2,y_i)}\right\}^2\right],$$

for $i = 1,\ldots,n+1$, where $c'(1/\sigma^2,y_i)$ and $c''(1/\sigma^2,y_i)$ are, respectively, the first and the second derivatives of $c(\lambda,y_i)$, with respect to λ, computed at $\lambda = 1/\sigma^2$. The above likelihood quantities are $i_{\sigma^2\sigma^2} = -E(\sum_{i=1}^n \ell_{i;\sigma^2\sigma^2})$, $\upsilon_{\sigma^2,tu} = E(\sum_{i=1}^n \ell_{i;\sigma^2}\sum_{i=1}^n \ell_{i;tu})$, $\upsilon_{\sigma^2,\sigma^2\sigma^2} = E(\sum_{i=1}^n \ell_{i;\sigma^2}\sum_{i=1}^n \ell_{i;\sigma^2\sigma^2})$, $\upsilon_{\sigma^2,t,u} = E(\sum_{i=1}^n \ell_{i;\sigma^2}\sum_{i=1}^n \ell_{i;t}\sum_{i=1}^n \ell_{i;u})$ and $\upsilon_{\sigma^2,\sigma^2,\sigma^2} = E(\sum_{i=1}^n \ell_{i;\sigma^2}\sum_{i=1}^n \ell_{i;\sigma^2}\sum_{i=1}^n \ell_{i;\sigma^2})$, where

$$\ell_{i;s} = \partial \log p_i(y_i;\omega,\lambda)/\partial\omega_s = \sigma^{-2}(y_i - \mu_i)\theta_{i,s},$$
$$\ell_{i;tu} = \partial^2 \log p_i(y_i;\omega,\lambda)/\partial\omega_t\partial\omega_u = -\sigma^{-2}\{\mu_{i,u}\theta_{i,t} - (y_i - \mu_i)\theta_{i,tu}\},$$

for $i = 1,\ldots,n$ and $s,t,u = 1,\ldots,d$. Since, as usual, $E(\ell_{i;s}) = 0$ and $E(\ell_{i;\sigma^2}) = 0$, the calculations are in fact simplified. When the order of these likelihood quantities is $O(n)$, then $\tilde{\sigma}^2_\alpha = \hat{\sigma}^2 + O(n^{-1})$ and the order of $H_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$ and $S_{n+1}(z;\tilde{\omega}_\alpha,\tilde{\lambda}_\alpha)$ is $O(n^{-1})$. These modifying terms can be evaluated at $(\omega,\sigma^2) = (\hat{\omega},\hat{\sigma}^2)$ without changing the order of approximation.

4. Examples

In this final section, generalized linear models with normal and gamma distributed error terms, which present explicit closed form solutions for the prediction problem, are taken into account. Models with an inverse Gaussian distributed error term do not allow analogous explicit solutions. The discrete case, which is of interest for a number of potential applications, requires particular attention. Here, we shall consider generalized linear models with a Poisson distributed error term, which are characterized by a computationally tractable predictive solution. Analogous results, not presented in this paper, may be obtained for models with a binomial distributed error term.

4.1 *Normal error*

Let us assume that $Y_1, \ldots, Y_n, Y_{n+1}$, with $Z = Y_{n+1}$, are mutually independent normally distributed random variables with mean μ_i, $i = 1, \ldots, n+1$, and known variance σ^2; namely, $N(\mu_i, \sigma^2)$. A generic link function $g(\cdot)$ is considered. Thus, Y_i, $i = 1, \ldots, n+1$, is a reproductive exponential dispersion model with $\mu_i = g^{-1}(x_i^T\omega)$, $V(\mu_i) = 1$, $\theta_i = \mu_i$ and $K(\theta_i) = \theta_i^2/2$. The first two orthogonal polynomials are, respectively,

$$B_1(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})}{\sigma}, \qquad B_2(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})^2}{\sigma^2} - 1$$

and the estimative prediction limit based on the maximum likelihood estimator $\hat{\omega}$ and that one based on the modified estimator $\tilde{\omega}_\alpha$ are, respectively, $\hat{z}_\alpha^e = \hat{\mu}_{n+1} + u_\alpha\sigma$ and $\tilde{z}_\alpha^e = \tilde{\mu}_{n+1} + u_\alpha\sigma$, with u_α the α-quantile of the standard normal distribution.

From (3.7), using properties of Hermite polynomials (see (A.1) and (A.2) in the Appendix), it is almost immediate to approximate the improved α-prediction limit as

$$\tilde{z}_\alpha^+ = \tilde{\mu}_{n+1} + \frac{1}{2}\sigma A_1(x_{n+1}; \tilde{\omega}_\alpha, \sigma) + u_\alpha\sigma\left\{1 - \frac{1}{2}A_2(x_{n+1}; \tilde{\omega}_\alpha, \sigma)\right\},$$

where $\tilde{\omega}_{\alpha r} = \hat{\omega}_r + u_\alpha\sigma x_{n+1,s}\{g'(\hat{\mu}_{n+1})\}^{-1}\hat{\Delta}^{rs}$, $r = 1, \ldots, d$. By (3.9), we obtain the alternative approximation

$$\hat{z}_\alpha^\circ = \hat{\mu}_{n+1} + \frac{1}{2}\sigma A_1(x_{n+1}; \hat{\omega}, \sigma) + u_\alpha\sigma\left\{1 + \frac{1}{2}A_2(x_{n+1}; \hat{\omega}, \sigma)\right\}. \tag{4.1}$$

Here

$$A_1(x_{n+1}; \omega, \sigma) = \frac{\sigma x_{n+1,r}x_{n+1,s}g''(\mu_{n+1})}{\{g'(\mu_{n+1})\}^3}\Delta^{rs} - \frac{\sigma x_{n+1,r}}{g'(\mu_{n+1})}\Sigma_{stu}\Delta^{tu}\Delta^{rs},$$
$$A_2(x_{n+1}; \omega, \sigma) = \frac{x_{n+1,r}x_{n+1,s}}{\{g'(\mu_{n+1})\}^2}\Delta^{rs},$$

with $\Sigma_{stu} = \sum_{i=1}^n g''(\mu_i)x_{i,s}x_{i,t}x_{i,u}\{g'(\mu_i)\}^{-4}$ and Δ^{rs} the (r,s)-element of the inverse of matrix $[\Delta_{rs}]$, with $\Delta_{rs} = \sum_{i=1}^n x_{i,r}x_{i,s}\{g'(\mu_i)\}^{-2}$. Retaining the same order of error, $\tilde{z}_\alpha^+$ may be further approximated by

$$\hat{\mu}_{n+1} + \frac{1}{2}\sigma A_1(x_{n+1}; \hat{\omega}, \sigma) + u_\alpha\sigma\sqrt{1 + A_2(x_{n+1}; \hat{\omega}, \sigma)}, \tag{4.2}$$

that is, by the α-quantile of a $N(\hat{\mu}_{n+1} + \frac{1}{2}\sigma A_1(x_{n+1}; \hat{\omega}, \sigma), \sigma^2\{1 + A_2(x_{n+1}; \hat{\omega}, \sigma)\})$ distribution. A straightforward interpretation of formula (4.2) is possible since, to third order accuracy, $-\frac{1}{2}\sigma A_1(x_{n+1}; \omega, \sigma)$ and $\sigma^2 A_2(x_{n+1}; \omega, \sigma)$ are, respectively, the bias and the variance of the estimator $\hat{\mu}_{n+1} = g^{-1}(x_{n+1}^T\hat{\omega})$. Note that, in the Gaussian case, the dependence of the modified predictive density (2.2) on the target value α is, in fact, suppressed to the relevant order of approximation. If the canonical link function $g(\mu) = \mu$ is considered, we have that $\mu_{n+1} = x_{n+1}^T\omega$, $A_1(x_{n+1}; \omega, \sigma) = 0$ and $A_2(x_{n+1}; \omega, \sigma) = x_{n+1,r}x_{n+1,s}\Delta^{rs}$, with $\Delta_{rs} = \sum_{i=1}^n x_{i,r}x_{i,s}$, and (4.2) coincides with the well-known prediction limit, which satisfies relation (2.1) exactly.

When σ is unknown, since

$$\ell_{i;\sigma^2} = \frac{1}{2\sigma^4}\{(y_i - \mu_i)^2 - \sigma^2\}, \qquad \ell_{i;\sigma^2\sigma^2} = \frac{1}{2\sigma^6}\{\sigma^2 - (y_i - \mu_i)^2\},$$

$i = 1, \ldots, n+1$, and $i_{\sigma^2\sigma^2} = n/(2\sigma^4)$, $v_{\sigma^2,tu} = 0$, $v_{\sigma^2,t,u} = \sigma^{-4}\Delta_{tu}$, $v_{\sigma^2,\sigma^2\sigma^2} = -n/\sigma^6$, $v_{\sigma^2,\sigma^2,\sigma^2} = n/\sigma^6$, the predictive density (3.10), which gives improved prediction limits, corresponds to

$$\phi(z; \tilde{\mu}_{n+1}, \tilde{\sigma}_\alpha)\left[1 + \frac{1}{2}\tilde{z}A_1(x_{n+1}; \tilde{\omega}_\alpha, \tilde{\sigma}_\alpha) - \frac{1}{2}(\tilde{z}^2 - 1)A_2(x_{n+1}; \tilde{\omega}_\alpha, \tilde{\sigma}_\alpha) - \frac{1}{4}\{\tilde{z}^4 - \tilde{z}^2(6 + 2d) + 3 + 2d\}n^{-1}\right],$$

where $A_1(\cdot)$, $A_2(\cdot)$, Δ_{tu} and $\tilde{\omega}_\alpha$ are given above, $\phi(\cdot; \mu, \sigma)$ is the density of a $N(\mu, \sigma^2)$ distribution and $\tilde{z} = \tilde{\sigma}_\alpha^{-1}(z - \tilde{\mu}_{n+1})$, with $\tilde{\sigma}_\alpha^2 = \hat{\sigma}^2 + \{(\hat{z}_\alpha^e - \hat{\mu}_{n+1})^2 - \hat{\sigma}^2\}n^{-1}$. Integration using properties of Hermite polynomials, followed by inversion and by a suitable Cornish-Fisher type expansion for u_α, shows that, in this case, the improved prediction limit $\tilde{z}_\alpha^+$ may be approximated by (4.1) and (4.2), with $\hat{\sigma}\sqrt{n/(n-d)}$ substituted for σ and t_α, the α-quantile of a Student t distribution with $n - d$ degrees of freedom, substituted for u_α. With the canonical link function, the second approximation coincides with the well-known prediction limit for linear regression models, which satisfies relation (2.1) exactly.

With regard to the calibration problem, an α-level confidence interval for x_{n+1} can be obtained by inversion of the associated α-prediction interval based on the improved prediction limit $z_\gamma^+(\tilde{\omega}_\gamma, \sigma; x_{n+1})$ or on the approximations (4.1) and (4.2). We consider the case with σ known, since the general case with σ unknown is obtained with the usual substitution for σ and u_α. In particular, if we consider the canonical link function $g(\mu) = \mu$, we obtain the exact prediction limit for normal linear regression models. For $\omega = (\omega_1, \omega_2)^T$ and $x_i = (1, x_{i,2})^T$, $i = 1, \ldots, n+1$, it is given by (4.2) with $\hat{\mu}_{n+1} = \hat{\omega}_1 + x_{n+1,2}\hat{\omega}_2$, $A_1(x_{n+1}; \hat{\omega}, \sigma) = 0$ and $A_2(x_{n+1}; \hat{\omega}, \sigma) = 1/n + (x_{n+1,2} - \bar{x})^2/S_{xx}$. Thus, in this case, the results on calibration are analogous to those reviewed by Brown ((1993), § 2.3).

A simple simulation study confirms the superiority of the calibration procedure based on the improved prediction limits over those based on an approximate pivotal quantity and on the estimative prediction limits. We consider a normal model with a logarithmic link function $g(\mu) = \log(\mu)$; indeed, we assume σ known and scalar ω and x_i, $i = 1, \ldots, n+1$. Thus, $\mu_{n+1} = \exp(\omega x_{n+1})$, $A_1(x_{n+1}; \omega, \sigma) = \sigma x_{n+1}\exp(\omega x_{n+1})\{\Sigma_{(3)}\Delta_{(2)}^{-1} - x_{n+1}\}\Delta_{(2)}^{-1}$ and $A_2(x_{n+1}; \omega, \sigma) = x_{n+1}^2\exp(2\omega x_{n+1})\Delta_{(2)}^{-1}$, with $\Delta_{(2)} = \sum_{i=1}^n x_i^2\exp(2\omega x_i)$ and $\Sigma_{(3)} = \sum_{i=1}^n x_i^3\exp(2\omega x_i)$. Moreover, we consider the approximate pivotal quantity $\{Z - \exp(\hat{\omega}x_{n+1})\}/(\sigma^2 + \hat{V}_g)^{1/2}$, which gives the approximate α-prediction limit $\hat{z}_\alpha^a = \hat{\omega}x_{n+1} + u_\alpha(\sigma^2 + \hat{V}_g)^{1/2}$, where $\hat{V}_g = \sigma^2 A_2(x_{n+1}; \hat{\omega}, \sigma)$ is an estimator for the variance of $\exp(\hat{\omega}x_{n+1})$ (see, for example, Seber and Wild (1988), § 5.12).

Sample of size $n = 10$ are generated from this model with $\omega = 0.15$ and $\sigma = 1$. The covariates x for the calibration stage are between 20 and 50, while, for $\xi = x_{n+1}$, we assume the three alternative values 35, 45 and 50. Table 1 gives estimates, with the corresponding estimated standard errors, of the actual confidence level for α-level calibration intervals, with $\alpha = 0.9, 0.95, 0.99$, based on $\hat{z}_\gamma^e$, $\hat{z}_\gamma^a$ and $\hat{z}_\gamma^\circ$. These simulations, based on 10,000 replications, show that for $\xi = 35$, which represent a good situation, these methods are equivalent; however, for $\xi = 50$, which represent a problematic situation, the method based on improved prediction intervals performs uniformly better. The method based on the estimative prediction intervals seem to be inadequate even in the intermediate situation with $\xi = 45$.

Table 1. Coverage probabilities, with estimated standard errors, for α-level calibration intervals based on prediction limits. Simulations from the normal model with $d = 1$, $g(\mu) = \log(\mu)$, $\sigma = 1$ known and $\omega = 0.15$. Covariates x between 20 and 50 and (a) $x_{n+1} = 35$, (b) $x_{n+1} = 45$, (c) $x_{n+1} = 50$. Estimation based on 10,000 replications of samples of size $n = 10$.

Prediction	$\alpha = 0.9$			$\alpha = 0.95$			$\alpha = 0.99$		
Limit	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Estimative	0.897	0.855	0.687	0.948	0.913	0.771	0.988	0.977	0.885
	0.003	0.003	0.005	0.002	0.003	0.004	0.001	0.001	0.003
Approximate	0.898	0.886	0.840	0.949	0.935	0.906	0.989	0.986	0.975
	0.003	0.003	0.004	0.002	0.002	0.003	0.001	0.001	0.002
Improved	0.898	0.888	0.860	0.949	0.936	0.923	0.989	0.987	0.982
	0.003	0.003	0.003	0.002	0.002	0.003	0.001	0.001	0.001

4.2 *Gamma error*

Let us assume that $Y_1, \ldots, Y_n, Y_{n+1}$, with $Z = Y_{n+1}$, are mutually independent gamma distributed random variables with scale parameter ϕ_i, $i = 1, \ldots, n+1$, and known shape parameter ν; namely, $Ga(\nu, \phi_i)$. For ease of exposition, the canonical link function $g(\mu) = -1/\mu$ is considered. Similar results may be obtained for a generic link function $g(\cdot)$. Thus, Y_i, $i = 1, \ldots, n+1$, is a reproductive exponential dispersion model with $\mu_i = \nu\phi_i = -(x_i^T\omega)^{-1}$, $V(\mu_i) = \mu_i^2$, $\theta_i = -1/\mu_i$, $\sigma^2 = 1/\nu$ and $K(\theta_i) = -\log(-\theta_i)$. The first two orthogonal polynomials are, respectively,

$$B_1(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})}{\sigma\mu_{n+1}}, \quad B_2(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})^2}{\sigma^2\mu_{n+1}^2} - 2\frac{(z - \mu_{n+1})}{\mu_{n+1}} - 1,$$

and the estimative prediction limit based on the maximum likelihood estimator $\hat{\omega}$ and that one based on the modified estimator $\tilde{\omega}_\alpha$ are, respectively, $\hat{z}_\alpha^e = u_\alpha\hat{\phi}_{n+1} = u_\alpha\hat{\mu}_{n+1}\sigma^2$ and $\tilde{z}_\alpha^e = u_\alpha\tilde{\phi}_{n+1} = u_\alpha\tilde{\mu}_{n+1}\sigma^2$, with u_α the α-quantile of a $Ga(\nu, 1)$ distribution.

From (3.7), using properties of Laguerre polynomials (see (A.3) and (A.4) in the Appendix), we obtain the following explicit approximation for the improved α-prediction limit

$$(4.3) \quad \tilde{z}_\alpha^+ = u_\alpha\tilde{\mu}_{n+1}\sigma^2\left\{1 + \frac{1}{2}\sigma A_1(x_{n+1}; \tilde{\omega}_\alpha, \sigma) - \frac{1}{2}(u_\alpha\sigma^2 - 1 + \sigma^2)A_2(x_{n+1}; \tilde{\omega}_\alpha, \sigma)\right\},$$

where $\tilde{\omega}_{\alpha r} = \hat{\omega}_r + (u_\alpha\sigma^2 - 1)x_{n+1,s}\hat{\mu}_{n+1}\hat{\Delta}^{rs}$, $r = 1, \ldots, d$, and, by (3.9), the alternative approximation

$$(4.4) \quad \hat{z}_\alpha^\circ = u_\alpha\hat{\mu}_{n+1}\sigma^2\left\{1 + \frac{1}{2}\sigma A_1(x_{n+1}; \hat{\omega}, \sigma) + \frac{1}{2}(u_\alpha\sigma^2 - 1 + \sigma^2)A_2(x_{n+1}; \hat{\omega}, \sigma)\right\}.$$

Here

$$A_1(x_{n+1}; \omega, \sigma) = -2\sigma x_{n+1,r}x_{n+1,s}\mu_{n+1}^2\Delta^{rs} - \sigma x_{n+1,r}\mu_{n+1}\Sigma_{stu}\Delta^{tu}\Delta^{rs},$$
$$A_2(x_{n+1}; \omega, \sigma) = x_{n+1,r}x_{n+1,s}\mu_{n+1}^2\Delta^{rs},$$

with Δ^{rs} the (r, s)-element of the inverse of matrix $[\Delta_{rs}]$, with $\Delta_{rs} = \sum_{i=1}^n x_{i,r}x_{i,s}\mu_i^2$, and $\Sigma_{stu} = -2\sum_{i=1}^n x_{i,s}x_{i,t}x_{i,u}\mu_i^3$. Retaining the same order of error, $\tilde{z}_\alpha^+$ may be approximated by the α-quantile of a $Ga(\nu_\alpha^\circ, \hat{\phi}_{n+1}^\circ)$ distribution, with $\nu_\alpha^\circ = \sigma^{-2}\{1 + \frac{1}{2}\sigma^2(u_\alpha +$

$1)A_2(x_{n+1};\hat{\omega},\sigma)+\frac{1}{2}\sigma A_1(x_{n+1};\hat{\omega},\sigma)\}^{-1}$ and $\hat{\phi}^{\circ}_{n+1}=\hat{\mu}_{n+1}\{1-\frac{1}{2}A_2(x_{n+1};\hat{\omega},\sigma)\}/\nu^{\circ}_{\alpha}$, provided that the $\nu^{\circ}_{\alpha},\hat{\phi}^{\circ}_{n+1}>0$. Note that, in the gamma case, the dependence of the modified predictive density (2.2) on the target value α maintains.

With regard to the calibration problem, an α-level confidence interval for x_{n+1} can be obtained by inversion of the associated improved α-prediction interval based on (4.3) or on the approximation (4.4). Here, we consider the simple case with scalar ω and x_i, $i=1,\ldots,n+1$. In this particular situation, the improved α-prediction limit is approximated, up to terms of order $O(n^{-1})$, by (4.4) with $\hat{\mu}_{n+1}=-1/(\hat{\omega}x_{n+1})$, $A_1(x_{n+1};\hat{\omega},\sigma)=0$ and $A_2(x_{n+1};\hat{\omega},\sigma)=1/n$; that is,

$$z^{\circ}_{\alpha}(\hat{\omega},\sigma;x_{n+1})=\frac{u_{\alpha}}{\nu\hat{\omega}x_{n+1}}\{1+(u_{\alpha}-\nu+1)/(2\nu n)\}.$$

Whenever $1+(u_{\gamma}-\nu+1)/(2\nu n)>0$, with $\gamma=(1-\alpha)/2,(1+\alpha)/2$, and $\hat{\omega}<0$, the associated confidence interval for x_{n+1} is $(x_{(1-\alpha)/2}(\hat{\omega},\sigma;z),x_{(1+\alpha)/2}(\hat{\omega},\sigma;z))$, where z is the observation available in the prediction stage and

$$x_{\gamma}(\hat{\omega},\sigma;z)=-\frac{u_{\gamma}}{\nu\hat{\omega}z}\{1+(u_{\gamma}-\nu+1)/(2\nu n)\}.$$

If $\hat{\omega}>0$, the confidence interval for x_{n+1} is obtained by reversing the extremes. With alternative link functions, an explicit expression for the confidence interval for x_{n+1} may not exists; however, it can be obtained by means of a suitable numeric inversion procedure.

A simple simulation study confirms the superiority, in terms of actual coverage level, of the calibration intervals based on the improved prediction limits over those based on the estimative prediction limits. We consider a gamma model with the canonical link function $g(\mu)=-1/\mu$; indeed, we assume ν known and scalar ω and x_i, $i=1,\ldots,n+1$. Sample of size $n=10$ are generated from this model with $\omega=-0.08$ and $\nu=2$. The covariates x for the calibration stage are between 2 and 12, while, for $\xi=x_{n+1}$, we assume the three alternative values 3, 7 and 11. Table 2 gives estimates of the actual confidence level, with the corresponding estimated standard errors, for α-level calibration intervals, with $\alpha=0.9,0.95,0.99$, based on $\hat{z}^{e}_{\gamma}$, and $\hat{z}^{\circ}_{\gamma}$. The simulations, based on 10,000 replications, show that the method based on the improved prediction intervals performs uniformly better.

Table 2. Coverage probabilities, with estimated standard errors, for α-level calibration intervals based on prediction limits. Simulations from the gamma model with $d=1$, $g(\mu)=-1/\mu$, $\nu=2$ known and $\omega=-0.08$. Covariates x between 2 and 12 and (a) $x_{n+1}=3$, (b) $x_{n+1}=7$, (c) $x_{n+1}=11$. Estimation based on 10,000 replications of samples of size $n=10$.

Prediction	$\alpha=0.9$			$\alpha=0.95$			$\alpha=0.99$		
Limit	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Estimative	0.881	0.877	0.883	0.939	0.937	0.935	0.986	0.984	0.982
	0.003	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.001
Improved	0.901	0.898	0.902	0.951	0.952	0.949	0.992	0.989	0.990
	0.003	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.001

4.3 *Poisson error*

Let us assume that $Y_1, \ldots, Y_n, Y_{n+1}$, with $Z = Y_{n+1}$, are mutually independent Poisson distributed random variables with mean μ_i, $i = 1, \ldots, n+1$; namely, $Po(\mu_i)$. For ease of exposition, the canonical link function $g(\mu) = \log(\mu)$ is considered, that is a log-linear model is defined. Similar results may be obtained for a generic link function $g(\cdot)$. Thus, Y_i, $i = 1, \ldots, n+1$, is a reproductive exponential dispersion model with $\mu_i = \exp(x_i^T \omega)$, $V(\mu_i) = \mu_i$, $\sigma^2 = 1$, $\theta_i = \log(\mu_i)$ and $K(\theta_i) = \exp(\theta_i)$. The first two orthogonal polynomials are

$$B_1(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})}{\sqrt{\mu_{n+1}}}, \tag{4.5}$$

$$B_2(z, x_{n+1}; \omega, \sigma) = \frac{(z - \mu_{n+1})^2}{\mu_{n+1}} - \frac{(z - \mu_{n+1})}{\mu_{n+1}} - 1,$$

and the estimative prediction limits based on the maximum likelihood estimator $\hat{\omega}$ and on the modified estimator $\tilde{\omega}_\alpha$ are, respectively, $\hat{z}_\alpha^e = \inf\{z \in I\!N : F_{n+1}(z; \hat{\omega}, \lambda) \geq \alpha\}$ and $\tilde{z}_\alpha^e = \inf\{z \in I\!N : F_{n+1}(z; \tilde{\omega}_\alpha, \lambda) \geq \alpha\}$, with $F_{n+1}(z; \hat{\omega}, \lambda)$ and $F_{n+1}(z; \tilde{\omega}_\alpha, \lambda)$ the corresponding estimative distribution functions.

From (3.6), particularized for discrete random variables, using formulae (A.5) and (A.6) given in the Appendix, we determine the distribution function associated to the predictive density (2.2), that is,

$$F_{n+1}^+(z; \tilde{\omega}_\alpha, \lambda) = F_{n+1}(z; \tilde{\omega}_\alpha, \lambda) + \frac{1}{2}\{(z^* - \tilde{\mu}_{n+1}) A_2(x_{n+1}; \tilde{\omega}_\alpha, \sigma) \tag{4.6}$$
$$- \sqrt{\tilde{\mu}_{n+1}} A_1(x_{n+1}; \tilde{\omega}_\alpha, \sigma)\} p_{n+1}(z^*; \tilde{\omega}_\alpha, \lambda),$$

where $\tilde{\omega}_{\alpha r} = \hat{\omega}_r + (\hat{z}_\alpha^e - \hat{\mu}_{n+1}) x_{n+1,s} \hat{\Delta}^{rs}$, $r = 1, \ldots, d$, and, by (3.8), its approximation

$$F_{n+1}^\circ(z; \hat{\omega}, \lambda, \alpha) \tag{4.7}$$
$$= F_{n+1}(z; \hat{\omega}, \lambda) + \frac{1}{2}\{(z^* - 2\hat{z}_\alpha^e + \hat{\mu}_{n+1}) A_2(x_{n+1}; \hat{\omega}, \sigma)$$
$$- \sqrt{\hat{\mu}_{n+1}} A_1(x_{n+1}; \hat{\omega}, \sigma)\} p_{n+1}(z^*; \hat{\omega}_\alpha, \lambda),$$

where $z^* = [z]$, namely, the integer part of z. Indeed,

$$A_1(x_{n+1}; \omega, \sigma) = -x_{n+1,r} x_{n+1,s} \sqrt{\mu_{n+1}} \Delta^{rs} - x_{n+1,r} \sqrt{\mu_{n+1}} \Sigma_{stu} \Delta^{tu} \Delta^{rs},$$
$$A_2(x_{n+1}; \omega, \sigma) = x_{n+1,r} x_{n+1,s} \mu_{n+1} \Delta^{rs},$$

with $\Sigma_{stu} = -\sum_{i=1}^n x_{i,s} x_{i,t} x_{i,u} \mu_i$ and Δ^{rs} the (r,s)-element of the inverse of matrix $[\Delta_{rs}]$, with $\Delta_{rs} = \sum_{i=1}^n x_{i,r} x_{i,s} \mu_i$. Thus, the improved α-prediction limit is defined as $\hat{z}_\alpha^+ = \inf\{z \in I\!N : F_{n+1}^+(z; \tilde{\omega}_\alpha, \lambda) \geq \alpha\}$, with $\hat{z}_\alpha^\circ = \inf\{z \in I\!N : F_{n+1}^\circ(z; \hat{\omega}, \lambda, \alpha) \geq \alpha\}$ its approximation based on (4.7). Note that, in the Poisson case, the dependence of the modified predictive distribution functions (4.6) and (4.7) on the target value α maintains.

With regard to the calibration problem, an α-level confidence region for x_{n+1} can be obtained by inversion of the associated improved α-prediction interval based on $z_\gamma^+(\hat{\omega}, \sigma; x_{n+1})$ or on its approximation $z_\gamma^\circ(\hat{\omega}, \sigma; x_{n+1})$. However, an explicit expression for the confidence region may not exist and it is usually obtained by means of a suitable numeric inversion procedure. The inversion requires special attention, since it may

result in a region different from an interval. Under this respect, investigation of the monotonicity of $F^{+}_{n+1}(z;\tilde{\omega}_\alpha,\lambda)$ or $F^{\circ}_{n+1}(z;\hat{\omega},\lambda,\alpha)$, as functions of x_{n+1} with z fixed, can be a useful preliminary analysis.

Besides the usual calibration regions for x_{n+1}, it may be of interest to estimate the value (or the set of values) for x_{n+1}, such that a fixed potential observation $z \in \mathbb{N}$ corresponds to the α-quantile of the future random variable Z. For example, given a suitable observation $z \in \mathbb{N}$, the set $\{x_{n+1} \in \mathbb{R} : z^{+}_{1/2}(\hat{\omega},\sigma;x_{n+1}) = z\}$ defines the covariate values for which z may be reasonably considered as the median observation in the prediction stage.

As noted by Vidoni (2001), the coverage probability of a prediction interval in the discrete case, and then the actual level of the associated confidence region for x_{n+1}, presents an unavoidable error component, due to discreteness. Indeed, the improved prediction limit z^{+}_{α} is superior to the estimative one, since it improves the actual coverage probability with regard to the second error component related to the estimative procedure. The same improvement maintains for the level of the associated confidence region for x_{n+1}. The following simulation study confirms this superiority.

Let us consider the simple case with scalar ω and x_i, $i = 1,\ldots,n+1$. In this particular situation, $F^{+}_{n+1}(z;\tilde{\omega}_\alpha,\lambda)$ and $F^{\circ}_{n+1}(z;\hat{\omega},\lambda,\alpha)$ are given by (4.6) and (4.7) with $\mu_{n+1} = \exp(x_{n+1}\omega)$ and

$$A_1(x_{n+1};\omega,\sigma) = -\frac{x^2_{n+1}\exp\left(\frac{1}{2}x_{n+1}\omega\right)}{\sum_{i=1}^{n} x_i^2\exp(x_i\omega)} + \frac{x_{n+1}\exp\left(\frac{1}{2}x_{n+1}\omega\right)}{\{\sum_{i=1}^{n} x_i^2\exp(x_i\omega)\}^2}\sum_{i=1}^{n} x_i^3\exp(x_i\omega),$$

$$A_2(x_{n+1};\omega,\sigma) = \frac{x^2_{n+1}\exp(x_{n+1}\omega)}{\sum_{i=1}^{n} x_i^2\exp(x_i\omega)}.$$

Sample of size $n = 10$ are generated from this model with $\omega = 0.15$; the covariates x for the calibration stage are between 1 and 10, while, for $\xi = x_{n+1}$, we assume the three alternative values 5.5, 7.5 and 9.5. The quantiles have been randomized to take care of the discreteness of the future random variable and to detect the actual differences in the coverage probability. Table 3 gives estimates of the coverage probabilities, with the corresponding estimated standard errors, for (randomized) α-prediction intervals, with $\alpha = 0.9, 0.95, 0.99$, based on $\hat{z}^{e}_{\gamma}$, and $\hat{z}^{\circ}_{\gamma}$. The simulations, based on 10,000 replications, show that the improved prediction intervals, and then the associated calibration regions, perform uniformly better than those based on the estimative procedure. In particular,

Table 3. Coverage probabilities, with estimated standard errors, for α-level calibration intervals based on (randomized) prediction limits. Simulations from the Poisson model with $d = 1$, $g(\mu) = \log(\mu)$ and $\omega = 0.15$. Covariates x between 1 and 10 and (a) $x_{n+1} = 5.5$, (b) $x_{n+1} = 7.5$, (c) $x_{n+1} = 9.5$. Estimation based on 10,000 replications of samples of size $n = 10$.

Prediction	$\alpha = 0.9$			$\alpha = 0.95$			$\alpha = 0.99$		
Limit	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Estimative	0.837	0.872	0.853	0.863	0.913	0.915	0.885	0.948	0.970
	0.004	0.003	0.003	0.003	0.003	0.002	0.003	0.002	0.002
Improved	0.844	0.889	0.889	0.868	0.922	0.941	0.886	0.952	0.974
	0.004	0.003	0.003	0.003	0.003	0.002	0.003	0.002	0.002

the improvement is evident for the more problematic situations when $x_{n+1} = 7.5$ or $x_{n+1} = 9.5$.

Appendix: Results on orthogonal polynomials

A.1 *Normal distribution*

Let us assume that the random variable Z follows a $N(\mu, \sigma^2)$ distribution with density $p(z; \mu, \sigma)$. The properties of Hermite polynomials (see Barndorff-Nielsen and Cox (1989), § 1.6) assure that

$$\int_{-\infty}^{z} (u-\mu)\sigma^{-1} p(u;\mu,\sigma)du = -\sigma p(z;\mu,\sigma), \tag{A.1}$$

$$\int_{-\infty}^{z} \{(u-\mu)^2\sigma^{-2} - 1\} p(u;\mu,\sigma)du = -(z-\mu)p(z;\mu,\sigma). \tag{A.2}$$

A.2 *Gamma distribution*

Let us assume that the random variable Z follows a $Ga(\nu, \phi)$ distribution, which can be viewed as a reproductive exponential dispersion model with $\mu = \nu\phi$, $V(\mu) = \mu^2$, $\theta = -1/\mu$, $\sigma^2 = 1/\nu$ and $K(\theta) = -\log(-\theta)$. Let $p(z; \mu, \sigma)$ denote the corresponding density function. The properties of Laguerre polynomials (see Barndorff-Nielsen and Cox (1989), § 1.6) assures that

$$\int_{0}^{z} (u-\mu)(\sigma\mu)^{-1} p(u;\mu,\sigma)du = -\sigma z p(z;\mu,\sigma), \tag{A.3}$$

$$\begin{aligned}\int_{0}^{z} \{(u-\mu)^2(\sigma\mu)^{-2} - 2(u-\mu)\mu^{-1} - 1\} p(u;\mu,\sigma)du \\ = -z\{(z-\mu)\mu^{-1} - \sigma^2\} p(z;\mu,\sigma).\end{aligned} \tag{A.4}$$

A.3 *Poisson distribution*

Let us assume that the random variable Z follows a $Po(\mu)$ distribution and that $p(z; \mu)$ denotes the corresponding probability function. It is easy to verify the following results on the first and the second order differences

$$\begin{aligned} p^{(1)}(z;\mu) &= p(z;\mu) - p(z-1;\mu) = -p(z;\mu)B_1(z;\mu)/\sqrt{\mu}, \qquad z \in I\!N, \\ p^{(2)}(z;\mu) &= p^{(1)}(z;\mu) - p^{(1)}(z-1;\mu) = p(z;\mu)B_2(z;\mu)/\mu, \qquad z \in I\!N, \end{aligned}$$

with the convention that $p(z; \mu) = 0$ for $z \notin I\!N$. Here, $B_1(z; \mu)$ and $B_2(z; \mu)$ correspond to (4.5), with μ substituted for μ_{n+1}. From the above relations, since

$$\sum_{u=0}^{z} p^{(1)}(u;\mu) = p(z;\mu), \qquad \sum_{u=0}^{z} p^{(2)}(u;\mu) = p^{(1)}(z;\mu), \quad z \in I\!N,$$

we obtain

$$\sum_{u=0}^{z} p(u;\mu)B_1(u;\mu) = -\sqrt{\mu}p(z;\mu), \tag{A.5}$$

$$\sum_{u=0}^{z} p(u;\mu)B_2(u;\mu) = -(z-\mu)p(z;\mu). \tag{A.6}$$

REFERENCES

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Barndorff-Nielsen, O. E. and Cox, D. R. (1996). Prediction and asymptotics, *Bernoulli*, **2**, 319–340.

Bellio, R. (2000). Likelihood methods for controlled calibration, Working Paper 3.2000, Department of Statistics, University of Udine Udine.

Brown, P. J. (1993). *Measurement, Regression and Calibration*, Clarendon Press, Oxford.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research, *Ann. Math. Statist.*, **10**, 162–186.

Gruet, M. A., Huet, S. and Jolivet, E. (1993). Practical use of bootstrap in regression, *Computer Intensive Methods in Statistics* (eds. W. Härdle and L. Simar), 150–166, Physica, Heidelberg.

Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman and Hall, London.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

Osborne, C. (1991). Statistical calibration: A review, *International Statistical Review*, **59**, 309–336.

Seber, G. A. and Wild, C. J. (1988). *Nonlinear Regression*, Wiley, New York.

Sundberg, R. (1999). Multivariate calibration-direct and indirect regression methodology (with discussion), *Scand. J. Statist.*, **26**, 161–207.

Vidoni, P. (1998). A note on modified estimative prediction limits and distributions, *Biometrika*, **85**, 949–953.

Vidoni, P. (2001). Improved prediction for continuous and discrete observations in generalized linear models, *Biometrika*, **88**, 881–887.