



UNIVERSITÀ
DEGLI STUDI
DI UDINE

Università degli studi di Udine

Calibrated multivariate distributions for improved conditional prediction

Original

Availability:

This version is available <http://hdl.handle.net/11390/1066937> since 2020-03-21T18:58:26Z

Publisher:

Published

DOI:10.1016/j.jmva.2015.08.001

Terms of use:

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

Calibrated multivariate distributions for improved conditional prediction

Paolo Vidoni

Department of Economics and Statistics, University of Udine

via Tomadini 30/a, I-33100 Udine, Italy

paolo.vidoni@uniud.it

Running headline: Calibrated multivariate distributions

Abstract

The specification of multivariate prediction regions, having coverage probability closed to the target nominal value, is a challenging problem both from the theoretical and the practical point of view. In this paper we define a well-calibrated multivariate predictive distribution giving suitable conditional prediction intervals with the desired overall coverage accuracy. This distribution is the extension in the multivariate setting of a calibrated predictive distribution defined for the univariate case and it is found on the idea of calibrating prediction regions for improving the coverage probability. This solution is asymptotically equivalent to that one based on asymptotic calculations and, whenever its explicit computation is not feasible, an approximation based on a simple bootstrap simulation procedure is readily available. Moreover, we state a simple, simulation-based, procedure for computing the associated improved conditional prediction limits.

Keywords: bootstrap calibration, coverage probability, prediction region, simultaneous prediction, time series.

1 Introduction

Predictive inference for an unobserved multivariate random variable may be of considerable interest in a number of application, such as the specification of simultaneous prediction intervals for future time series observations and the construction of prediction regions for observables from multivariate models. In this paper, prediction is considered from the frequentist perspective and the aim is to define a well-calibrated multivariate predictive distribution giving conditional prediction intervals, and in particular conditional prediction limits, with overall coverage probability closed to the target nominal value. The associated prediction regions are not necessarily of rectangular form, in the two-sided case, or of semi-infinite rectangular form, in the one-sided case. This is an important point and it will be appropriately discuss throughout

the paper.

Let (Y, Z) be a continuous random vector having joint density function $p(y, z; \theta)$, with $\theta \in \Theta \subseteq \mathbf{R}^d$, $d \geq 1$, an unknown d -dimensional parameter; $Y = (Y_1, \dots, Y_n)$, $n \geq 1$, is observable, while $Z = (Z_1, \dots, Z_m)$, $m \geq 1$, denotes a future, or yet unobserved, random vector. This is a fairly general formulation which includes both the simple case with Y_1, \dots, Y_n and Z independent, identically distributed, multivariate random variables and the more general situation with dependent Y and Z and, in particular, with Y and Z defined within a stochastic process model. For ease of exposition we consider $Y = (Y_1, \dots, Y_n)$ as a random vector with density $f(\cdot; \theta)$ and Z as an independent future random vector with density $g(\cdot; \theta)$, possibly different from $f(\cdot; \theta)$, with θ the same d -dimensional parameter as before; $G(\cdot; \theta)$ indicates the distribution function of Z . We also assume that $f(\cdot; \theta)$, $g(\cdot; \theta)$ and $G(\cdot; \theta)$ are sufficiently smooth functions of the parameter θ . The extension of the results to the case with dependent Y and Z is considered in the final part of the paper.

Although prediction problems may be tackled with different objectives, the aim here is to define an α -prediction region for Z , that is a random set $R(Y, \alpha) \subset \mathbf{R}^m$, depending on the observable sample Y and on the nominal coverage probability α , such that

$$P_{Y,Z}\{Z \in R(Y, \alpha); \theta\} = \alpha, \quad (1.1)$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0, 1)$. The above probability is called coverage probability and it is calculated with respect to the joint distribution of (Z, Y) ; moreover, it can be rewritten as $E_Y[P_Z\{Z \in R(Y, \alpha); \theta\}; \theta]$, where the expectation is with respect to Y and $P_Z\{\cdot; \theta\}$ is the probability distribution for Z .

Given a suitable predictive probability distribution for Z , namely $\tilde{P}_Z\{\cdot; Y\}$, defined as an

estimator for the true $P_Z\{\cdot; \theta\}$ based on the sample Y , we may define an α -prediction region for Z as a set $\tilde{R}(Y, \alpha) \subset \mathbf{R}^m$ such that

$$\tilde{P}_Z\{Z \in \tilde{R}(Y, \alpha); Y\} = \alpha.$$

Under this respect, we aim at introducing a predictive distribution function such that the corresponding α -prediction region $\tilde{R}(Y, \alpha)$ fulfils (1.1) exactly or with a high degree of accuracy, for each $\alpha \in (0, 1)$.

Although there are some special cases where there is an exact solution to (1.1), these situations are extremely rare. Thus, in general, we look for approximate solutions satisfying (1.1) almost exactly, for each $\alpha \in (0, 1)$. The easiest way for making prediction on Z is by using the estimative predictive distribution $P_Z\{\cdot; \hat{\theta}\}$, where the unknown parameter θ is substituted with an asymptotically efficient estimator $\hat{\theta}$ based on Y , such that $\hat{\theta} - \theta = O_p(n^{-1/2})$; we usually consider the maximum likelihood estimator or any asymptotically equivalent alternative estimator. However, it is well-known that the estimative α -prediction regions $R_e(Y, \alpha)$ are not entirely adequate predictive solutions, since their coverage probability differs from α by a term usually of order $O(n^{-1})$ and prediction statements may be rather inaccurate for small n . In fact, this naive solution underestimates the additional uncertainty introduced by assuming $\theta = \hat{\theta}$.

Concerning the univariate case, Barndorff-Nielsen and Cox (1994, 1996), Ueki and Fueda (2007) and Vidoni (1998, 2009) suggest a way to correct, by means of asymptotic calculations, the quantiles of the estimative predictive distribution, thus obtaining prediction intervals with a coverage error of order $o(n^{-1})$. A calibrating approach, useful in the multivariate case as well, has been suggested by Beran (1990) and applied, for example, by Hall *et al.* (1999), using a bootstrap procedure for improving the estimative prediction intervals. The key idea, behind

this approach, is to determine a suitable value $\bar{\alpha}$ such that the coverage probability of the estimative, recalibrated, prediction region $R_e(Y, \bar{\alpha})$ is equal or close to the target value α . The effect of (bootstrap) calibration is that of reducing the magnitude of the coverage error but it is valid for a specific α and it does not provide a general solution to the problem, such as those based based on the notion of predictive distribution.

Recently, Fonseca *et al.* (2014) extend the calibrating procedure to predictive distribution functions in such a way that the associated prediction intervals have coverage probability equal or close to the target value. This solution has similarities with that one specified by Lawless and Fredette (2005), involving (approximate) pivotal quantities, but it has the advantage that, whenever its computation is not feasible, it can be approximated using a suitable bootstrap simulation procedure or considering high-order asymptotic expansions, giving the same improved predictive distributions already known in the literature.

In the present paper, this result is properly extended to deal with multivariate prediction problems. In particular, a well-calibrated multivariate predictive distribution is derived. We prove that the associated joint predictive density is asymptotically equivalent to that one proposed by Corcuera and Giummolè (2006), which gives improved conditional prediction limits. This new solution, contrary to the Corcuera and Giummolè's one, has a simple and intuitive form and, when computations are hard to perform, it is readily available an approximation based on bootstrap simulation methods. Furthermore, generalizing a result presented in Ueki and Fueda (2007), we state a simple, simulation-based, procedure for computing the associated improved conditional prediction limits. The paper is organized as follows. Section 2 reviews some known results on improved predictive procedures. Section 3 introduces the new approach

based on the calibrated multivariate predictive distribution. Section 4 presents the procedure for calculating the improved conditional prediction limits, following the Ueki and Fueda's approach, and Section 5 briefly considers the extension to the case of dependent observations. Finally, Section 6 is dedicated to a simple example concerning autoregressive time series models.

2 Review on improved multivariate prediction

Let us review the calibrated predictive distribution function proposed by Fonseca *et al.* (2014) for the univariate case, that is for $m = 1$. Let us consider the estimative α -prediction limit $\hat{q}(\alpha) = q(\alpha; \hat{\theta})$, defined as the α -quantile of the estimative distribution function $G(z; \hat{\theta})$, such that $G\{\hat{q}(\alpha); \hat{\theta}\} = \alpha$. The associated coverage probability is

$$P_{Y,Z}\{Z \leq \hat{q}(\alpha); \theta\} = E_Y[G\{\hat{q}(\alpha); \theta\}; \theta] = C(\alpha; \theta).$$

By substituting α with $G(z; \hat{\theta})$ in $C(\alpha; \theta)$, we obtain

$$G_c(z; \hat{\theta}, \theta) = C\{G(z; \hat{\theta}); \theta\},$$

which is a predictive distribution function if $C(\cdot; \theta)$ is a sufficiently smooth function. The associated density function is a suitable modification of the estimative predictive density $g(z; \hat{\theta})$ given by

$$g_c(z; \hat{\theta}, \theta) = g(z; \hat{\theta})C'\{G(z; \hat{\theta}); \theta\},$$

with $C'(\alpha; \theta) = dC(\alpha; \theta)/d\alpha$. It is quite easy to prove that it gives, as quantiles, well-calibrated prediction limits $q_c(\alpha; \hat{\theta}, \theta)$ achieving, for all $\alpha \in (0, 1)$, coverage probability equal to the target nominal value α . Although $G_c(z; \hat{\theta}, \theta)$ depends on the unknown θ , and therefore it is not

useful in practice, we may consider the corresponding plug-in estimator $\hat{G}_c(z; \hat{\theta}) = G_c(z; \hat{\theta}, \hat{\theta})$ or suitable approximations based on asymptotic expansions or bootstrap simulation procedures, which give improved prediction limits having coverage error reduced to order $o(n^{-1})$.

Let us consider the multivariate setting assuming $m > 1$. We focus on a particular estimative prediction region $R_e(Y, \alpha)$ based on the system of prediction limits defined as quantiles of the conditional distributions of the components of vector $Z = (Z_1, \dots, Z_m)$. More precisely, we set

$$R_e(Y, \alpha) = \{z \in \mathbf{R}^m : z_i \leq \hat{q}_i(\alpha_i), i = 1, \dots, m\}, \quad (2.1)$$

where $\hat{q}_i(\alpha_i) = q_i(\alpha_i, z_{(i-1)}; \hat{\theta})$, $i = 1, \dots, m$, is the α_i -quantile of the conditional distribution of Z_i given $Z_{(i-1)} = z_{(i-1)}$, evaluated at $\theta = \hat{\theta}$, with $\hat{q}_1(\alpha_1) = q_1(\alpha_1; \hat{\theta}) = q_1(\alpha_1, z_{(0)}; \hat{\theta})$ the α_1 -quantile of the marginal distribution of Z_1 . Hereafter, we state $Z_{(i-1)} = (Z_1, \dots, Z_{i-1})$, and analogously for $z_{(i-1)}$; whenever $i = 1$, the conditional event $Z_{(i-1)} = z_{(i-1)}$ is not considered. Finally, we assume $\prod_{i=1}^m \alpha_i = \alpha$ in order to assure that $R_e(Y, \alpha)$ is an α -prediction region, namely that $P_Z\{Z \in R_e(Y, \alpha); \hat{\theta}\} = \alpha$. Note that the conditional prediction limit $\hat{q}_i(\alpha_i)$, for each $i = 2, \dots, m$, is obtained recursively as a function of the previous, unknown future observations $z_{(i-1)}$. Thus, the specification of the prediction region is strongly dependent on the factorization of the joint distribution of vector Z which is adopted. Although this fact may reduce the generality of the approach, there are situations where there exists a convenient ordering among the components of Z , such as within time series models or panel data models.

Furthermore, the choice of these particular prediction limits determines a shape for the multivariate estimative prediction region which is usually non-rectangular. This could be interpreted as a drawback, since in the applications it is quite common to specify a multivariate prediction region as a sequence of simultaneous marginal prediction intervals, associated to the

components of the future random vector Z . However, when one uses a rectangular prediction region, it is implicitly assumed that the first realizations of the future path do not influence the location and the size of the prediction intervals associated to the future subsequent observations. The conditional prediction regions studied in this paper overtake this limitation. Indeed, they enable a suitable conditional predictive analysis, where alternative sequences of conditional prediction intervals may be derived, according to different potential scenarios for the future observations. A related point concerns the visualization of the prediction region, which is surely possible for $m \leq 3$. Although, for larger values of m a complete graphical representation is unfeasible, we may consider as well a sequence of m conditional prediction intervals or prediction limits, specified by assuming a particular path forecast scenario for the future realization z . For example, as a path forecast we can consider a vector of point predictors for Z , defined according to a suitable predictive optimality criterion.

The coverage probability of the estimative prediction region (2.1) is

$$\begin{aligned} P_{Y,Z}\{Z \in R_e(Y, \alpha); \theta\} &= E_Y[P_Z\{Z \in R_e(Y, \alpha); \theta\}; \theta] = E_Y\left\{\int_{-\infty}^{\hat{q}_1(\alpha_1)} \cdots \int_{-\infty}^{\hat{q}_m(\alpha_m)} g(z; \theta) dz; \theta\right\} \\ &= C_m(\alpha_1, \dots, \alpha_m; \theta). \end{aligned} \quad (2.2)$$

Since $C_m(\alpha_1, \dots, \alpha_m; \theta) = \alpha + O(n^{-1})$, as $n \rightarrow +\infty$, there is a coverage error term of order $O(n^{-1})$, which may be not negligible, thus reducing the accuracy of the predictive procedure. Corcuera and Giummolè (2006) derive an explicit expression for the $O(n^{-1})$ error term and introduce a system of modified estimative prediction limits achieving a coverage error reduced to order $o(n^{-1})$. The modification consists of two quantities: the first one takes into account the additional uncertainty introduced by putting $\theta = \hat{\theta}$, and it corresponds to that one proposed by Barndorff-Nielsen and Cox (1994, 1996) and Vidoni (1998) in the univariate case, while the

second one considers the additional dependency introduced among the components of Z , since the α_i -quantiles $q_i(\alpha_i) = q_i(\alpha_i, z_{(i-1)}; \theta)$ of the conditional distribution of Z_i given $Z_{(i-1)} = z_{(i-1)}$, $i = 1, \dots, m$, are all estimated by substituting θ with the same $\hat{\theta}$. As the estimative one, also the Corcuera and Giummolè's prediction region does not present a rectangular shape.

Furthermore, the predictive density, which gives the modified prediction limits as (conditional) α_i -quantiles, $i = 1, \dots, m$, is defined as

$$\begin{aligned} \tilde{g}(z; Y) = & g(z; \hat{\theta}) \left[1 - b_r(\hat{\theta}) \log \hat{g}_r + \frac{1}{2} i^{rs}(\hat{\theta}) \left\{ \log \hat{g}_r \log \hat{g}_s - \log \hat{g}_{rs} + [2] \sum_{i=1}^m \hat{h}_i \right. \right. \\ & \left. \left. + [2] \sum_{i=2}^m \sum_{j < i} \hat{h}_{ij} \right\} \right], \end{aligned} \quad (2.3)$$

where $b_r(\hat{\theta})$, $r = 1, \dots, d$, is the $O(n^{-1})$ bias term $b_r(\theta)$ of the r -th component of the maximum likelihood estimator, evaluated at $\theta = \hat{\theta}$, and $i^{rs}(\hat{\theta})$, $r, s = 1, \dots, d$, is the (r, s) -element $i^{rs}(\theta)$ of the inverse of the expected information matrix based on Y , evaluated at $\theta = \hat{\theta}$. Moreover, $\log \hat{g}_r$ and $\log \hat{g}_{rs}$ are the first and the second partial derivatives of $\log g = \log g(z; \theta)$, with respect to the corresponding components of vector θ , evaluated at $\theta = \hat{\theta}$. Here we use index notation and the Einstein summation convention, according to which if an index occurs more than once in a summand then summation over that index is understood; $[2]$ indicates the sum of two terms obtained by permutation of the indices. We shall indicate with $G^i = G^i(z_i | z_{(i-1)}; \theta)$ and $g^i = g^i(z_i | z_{(i-1)}; \theta)$, $i = 1, \dots, m$, respectively, the distribution function and density function of Z_i given $Z_{(i-1)} = z_{(i-1)}$; if $i = 1$, we have $G^1 = G^1(z_1; \theta)$ and $g^1 = g^1(z_1; \theta)$, corresponding to the marginal distribution of Z_1 . Indeed,

$$\begin{aligned} \hat{h}_i &= h_i(z_{(i)}; \hat{\theta}) = \frac{(d \log \hat{g}_r^i / dz_i) \hat{G}_s^i}{\hat{g}^i}, \quad i = 1, \dots, m, \\ \hat{h}_{ij} &= h_{ij}(z_{(i)}; \hat{\theta}) = \left\{ \frac{d \hat{g}^i / dz_j}{\hat{g}^i} + \log \hat{g}_r^i \left(\frac{d}{dz_j} \log \prod_{k=j+1}^{i-1} \hat{g}^k \right) \right\} \frac{\hat{G}_s^j}{\hat{g}^j}, \quad i = 2, \dots, m, j < i, \end{aligned}$$

where $\log \hat{g}_r^i$ and \hat{G}_s^i are the first partial derivatives of $\log g^i$ and G^j , with respect to the corresponding components of vector θ , evaluated at $\theta = \hat{\theta}$. In the particular case where the components of Z are independent, the computation of the predictive density $\tilde{g}(z; Y)$ is greatly simplified.

3 Calibrating multivariate predictive distributions

The Corcuera and Giummolè's solution, thought theoretically notable, may be difficult to apply since the computation of the predictive density (2.3) and of the associated conditional prediction limits can be rather involved, and sometimes unfeasible, excluding the simple and not particularly interesting case of m independent future observations Z_1, \dots, Z_m . In this paper we shall overcome these practical difficulties and we shall specify a simpler, asymptotically equivalent predictive procedure, which can be actually considered in the multivariate setting in order to improve the estimative solution. This objective is achieved by defining a well-calibrated multivariate predictive distribution which is the multidimensional extension of that one introduced by Fonseca *et al.* (2014) in the univariate case and reviewed in Section 2.

Let us consider the estimative prediction region $R_e(Y, \alpha)$ defined by (2.1) and based on the system of estimative prediction limits $\hat{q}_i(\alpha_i)$, $i = 1, \dots, m$; the coverage probability of $R_e(Y, \alpha)$ corresponds to $C_m(\alpha_1, \dots, \alpha_m; \theta)$. By substituting α_i with $G^i(z_i | z_{(i-1)}; \hat{\theta})$ in $C_m(\alpha_1, \dots, \alpha_m; \theta)$, for each $i = 1, \dots, m$, we obtain

$$G_c(z; \hat{\theta}, \theta) = C_m\{G^1(z_1; \hat{\theta}), \dots, G^m(z_m | z_{(m-1)}; \hat{\theta}); \theta\}, \quad (3.1)$$

which is a proper multivariate predictive distribution function, provided that function $C_m(\cdot; \theta)$ is sufficiently smooth. The corresponding predictive density is the following modification of the

estimative predictive density $g(z; \hat{\theta})$

$$g_c(z; \hat{\theta}, \theta) = g(z; \hat{\theta}) \partial_1^m C_m \{G^1(z_1; \hat{\theta}), \dots, G^m(z_m | z_{(m-1)}; \hat{\theta}); \theta\}, \quad (3.2)$$

where ∂_1^m indicates partial differentiation of $C_m(\alpha_1, \dots, \alpha_m; \theta)$ with respect to $\alpha_1, \dots, \alpha_m$. Since the coverage probability $C_m(\alpha_1, \dots, \alpha_m; \theta)$ is defined by (2.2), provided that expectation with respect to Y and differentiation may be interchanged, we obtain that

$$\begin{aligned} \partial_1^m C_m(\alpha_1, \dots, \alpha_m; \theta) &= E_Y \left\{ \frac{\partial^m}{\partial \alpha_1 \dots \partial \alpha_m} \int_{-\infty}^{\hat{q}_1(\alpha_1)} \dots \int_{-\infty}^{\hat{q}_m(\alpha_m)} g(z; \theta) dz; \theta \right\} \\ &= E_Y \left[\frac{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_1; \hat{\theta}), \dots, q_m(\alpha_m, \hat{q}_{(m-1)}; \hat{\theta}); \theta\}}{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_1; \hat{\theta}), \dots, q_m(\alpha_m, \hat{q}_{(m-1)}; \hat{\theta}); \hat{\theta}\}}; \theta \right], \end{aligned}$$

where vectors $\hat{q}_{(1)}, \dots, \hat{q}_{(m-1)}$ are defined recursively by considering the updating relation $\hat{q}_{(i)} = (q_i(\alpha_i, \hat{q}_{(i-1)}; \hat{\theta}), \hat{q}_{(i-1)})$, for $i = 2, \dots, m-1$, with $\hat{q}_{(1)} = \hat{q}_1(\alpha_1)$.

Although $G_c(z; \hat{\theta}, \theta)$ and $g_c(z; \hat{\theta}, \theta)$ are not of direct use for prediction, since the true parameter value θ is unknown, we may determine, as in the univariate case, the corresponding plug-in estimators. We obtain the following result.

Proposition 3.1 *Under suitable regularity assumptions, assuring \sqrt{n} -consistency for the maximum likelihood estimator $\hat{\theta}$, the plug-in estimator $g_c(z; \hat{\theta}, \hat{\theta})$ of function (3.2) is such that*

$$g_c(z; \hat{\theta}, \hat{\theta}) = \tilde{g}(z; Y) + o_p(n^{-1}),$$

with $\tilde{g}(z; Y)$ given by (2.3).

PROOF See Appendix A.

Thus, function $g_c(z; \hat{\theta}, \hat{\theta})$ coincides, up to terms of order $O_p(n^{-1})$, to the predictive density defined by Corcuera and Giummolè (2006) and recalled in Section 2. For this reason, we can state that the new multivariate predictive distribution (3.1) evaluated at $\theta = \hat{\theta}$, namely

$G_c(z; \hat{\theta}, \hat{\theta})$, gives, as quantiles of the associated conditional distributions, prediction limits which are asymptotically equivalent to those ones specified by Corcuera and Giummolè (2006), and thus achieving a coverage probability equal to $\alpha + o(n^{-1})$. As a matter of fact, we presume that the improvement over the pure estimative solution could be substantial or even such that, as proved by Fonseca *et al.* (2014) in the univariate case, the target nominal value α is exactly achieved. Besides coverage accuracy, a further advantage in using the calibrated multivariate predictive distribution is that whenever explicit calculations or asymptotic approximations are not feasible, a relatively simple bootstrap simulation procedure is readily available for estimating the associated prediction limits.

To derive the conditional prediction limits associated to the multivariate calibrated predictive distribution $G_c(z; \hat{\theta}, \theta)$, we need a procedure for calculating the quantiles of the predictive distribution of Z_1 and of Z_i given $Z_{(i-1)} = z_{(i-1)}$, $i = 2, \dots, m$. From the multivariate joint predictive density (3.2), we get the corresponding marginal and conditional predictive densities

$$\begin{aligned} g_c^1(z_1; \hat{\theta}, \theta) &= g^1(z_1; \hat{\theta}) \partial_1 C_1\{G^1(z_1; \hat{\theta}); \theta\}, \\ g_c^i(z_i | z_{(i-1)}; \hat{\theta}, \theta) &= g^i(z_i | z_{(i-1)}; \hat{\theta}) \frac{\partial_1^i C_i\{G^1(z_1; \hat{\theta}), \dots, G^i(z_i | z_{(i-1)}; \hat{\theta}); \theta\}}{\partial_1^{i-1} C_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1} | z_{(i-2)}; \hat{\theta}); \theta\}}, \end{aligned}$$

with $i = 2, \dots, m$. Here,

$$C_j(\alpha_1, \dots, \alpha_j; \theta) = E_Y \left\{ \int_{-\infty}^{\hat{q}_1(\alpha_1)} \cdots \int_{-\infty}^{\hat{q}_j(\alpha_j)} g(z_1, \dots, z_j; \theta) dz_1 \cdots dz_j; \theta \right\},$$

for $j = 1, \dots, m$, is the coverage probability of the system of estimative prediction limits

$\hat{q}_1(\alpha_1), \dots, \hat{q}_j(\alpha_j)$ and ∂_1^j indicates partial differentiation with respect to $\alpha_1, \dots, \alpha_j$, so that

$$\partial_1^j C_j(\alpha_1, \dots, \alpha_j; \theta) = E_Y \left[\frac{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_1(\alpha_1); \hat{\theta}), \dots, q_j(\alpha_j, \hat{q}_{(j-1)}; \hat{\theta}); \theta\}}{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_1(\alpha_1); \hat{\theta}), \dots, q_j(\alpha_j, \hat{q}_{(j-1)}; \hat{\theta}); \hat{\theta}\}}; \theta \right].$$

It is quite easy to specify the associated marginal and conditional predictive distribution functions as

$$G_c^1(z_1; \hat{\theta}, \theta) = C_1\{G^1(z_1; \hat{\theta}); \theta\}, \quad (3.3)$$

$$G_c^i(z_i|z_{(i-1)}; \hat{\theta}, \theta) = \frac{\partial_1^{i-1} C_i\{G^1(z_1; \hat{\theta}), \dots, G^i(z_i|z_{(i-1)}; \hat{\theta}); \theta\}}{\partial_1^{i-1} C_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}); \theta\}}, \quad (3.4)$$

for $i = 2, \dots, m$, with $\partial_1^{i-1} C_i(\alpha_1, \dots, \alpha_i; \theta)$ given by

$$E_Y \left[\frac{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_{(1)}; \hat{\theta}), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}; \hat{\theta}); \theta\} G^i\{q_i(\alpha_i, \hat{q}_{(i-1)}; \hat{\theta})|\hat{q}_{(i-1)}; \theta\}}{g\{\hat{q}_1(\alpha_1), q_2(\alpha_2, \hat{q}_{(1)}; \hat{\theta}), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}; \hat{\theta}); \hat{\theta}\}}; \theta \right].$$

The quantiles $q_i^c(\alpha_i) = q_i^c(\alpha_i, z_{(i-1)}; \hat{\theta}, \theta)$, $i = 1, \dots, m$, which specify the system of improved prediction limits, are such that

$$G_c^1\{q_1^c(\alpha_1); \hat{\theta}, \theta\} = \alpha_1, \quad G_c^i\{q_i^c(\alpha_i)|z_{(i-1)}; \hat{\theta}, \theta\} = \alpha_i, \quad i = 2, \dots, m. \quad (3.5)$$

The plug-in estimators of these prediction limits, namely $\hat{q}_i^c(\alpha_i) = q_i^c(\alpha_i, z_{(i-1)}; \hat{\theta}, \hat{\theta})$, $i = 1, \dots, m$, define a system of prediction limits which is asymptotically equivalent to the Corcuera and Giummolè's (2006) solution and thus improves the estimative one.

As mentioned before, there is a fairly natural parametric bootstrap simulation procedure for estimating the predictive distribution functions $G_c^i(z_i|z_{(i-1)}; \hat{\theta}, \theta)$, $i = 1, \dots, m$, to be considered when an explicit expression for $C_i(\alpha_1, \dots, \alpha_i; \theta)$, $i = 1, \dots, m$, is not available. This is also a valid alternative to the improved multivariate predictive distribution proposed by Corcuera and Giummolè (2006), which is not very useful in applications since its computation can be very hard.

Let y_j^* , $j = 1, \dots, B$, be parametric bootstrap samples generated from $f(y; \hat{\theta})$ and let $\hat{\theta}_j^*$, $j = 1, \dots, B$, be the corresponding maximum likelihood estimates. Since $\partial_1^{i-1} C_{i-1}(\alpha_1, \dots, \alpha_{i-1}; \theta)$

and $\partial_1^{i-1}C_i(\alpha_1, \dots, \alpha_i; \theta)$ are defined as expectations with respect to Y , we define the associated bootstrap estimators, respectively, as

$$\partial_1^{i-1}\hat{C}_{i-1}(\alpha_1, \dots, \alpha_{i-1}; \theta)_b = \frac{1}{B} \sum_{j=1}^B \frac{g\{q_1(\alpha_1; \hat{\theta}_j^*), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}^j; \hat{\theta}_j^*); \hat{\theta}\}}{g\{q_1(\alpha_1; \hat{\theta}_j^*), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}^j; \hat{\theta}_j^*); \hat{\theta}_j^*\}},$$

$$\partial_1^{i-1}\hat{C}_i(\alpha_1, \dots, \alpha_i; \theta)_b = \frac{1}{B} \sum_{j=1}^B \frac{g\{q_1(\alpha_1; \hat{\theta}_j^*), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}^j; \hat{\theta}_j^*); \hat{\theta}\} G^i\{q_i(\alpha_i, \hat{q}_{(i-1)}^j; \hat{\theta}_j^*) | \hat{q}_{(i-1)}^j; \hat{\theta}\}}{g\{q_1(\alpha_1; \hat{\theta}_j^*), \dots, q_{i-1}(\alpha_{i-1}, \hat{q}_{(i-2)}^j; \hat{\theta}_j^*); \hat{\theta}_j^*\}},$$

where vectors $\hat{q}_{(1)}^j, \dots, \hat{q}_{(m-1)}^j$ are defined recursively by $\hat{q}_{(i)}^j = (q_i(\alpha_i, \hat{q}_{(i-1)}^j; \hat{\theta}_j^*), \hat{q}_{(i-1)}^j)$, for $i = 2, \dots, m-1$, with $\hat{q}_{(1)}^j = q_1(\alpha_1; \hat{\theta}_j^*)$. When $i = 1$, we have $\partial_1^{i-1}C_i(\alpha_1, \dots, \alpha_i; \theta) = C_1(\alpha_1; \theta)$, estimated by $\hat{C}_1(\alpha_1; \theta)_b = (1/B) \sum_{j=1}^B G^1\{q_1(\alpha_1; \hat{\theta}_j^*); \hat{\theta}\}$. Thus, the bootstrap estimators for the marginal and the conditional predictive distribution functions are

$$\hat{G}_c^1(z_1; \hat{\theta})_b = \hat{C}_1\{G^1(z_1; \hat{\theta}); \theta\}_b,$$

$$\hat{G}_c^i(z_i | z_{(i-1)}; \hat{\theta})_b = \frac{\partial_1^{i-1}\hat{C}_i\{G^1(z_1; \hat{\theta}), \dots, G^i(z_i | z_{(i-1)}; \hat{\theta}); \theta\}_b}{\partial_1^{i-1}\hat{C}_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1} | z_{(i-2)}; \hat{\theta}); \theta\}_b},$$

for $i = 2, \dots, m$. The corresponding α_i -quantiles, $i = 1, \dots, m$, define a system of prediction limits having an overall coverage probability equal to the target value α , with an error term which depends on the efficiency of the bootstrap simulation procedure.

4 Computing the improved prediction limits

The procedure introduced in the preceding section, for computing the calibrated conditional prediction limits or the associated plug-in estimators, is not direct and it involves the calculation of the joint predictive distribution (3.1), and in particular the associated conditional predictive distribution functions. Generalizing a result presented in Ueki and Fueda (2007), we obtain, as a useful alternative, a simple direct adjustment for the estimative prediction limits, which

is again equivalent to the system of improved prediction limits proposed by Corcuera and Giummolè (2006) in order to reduce the coverage error to order $o(n^{-1})$.

Let us consider the calibrated predictive distribution functions (3.3) and (3.4), calculated in the estimative prediction limits and evaluated at $\theta = \hat{\theta}$, namely

$$G_c^1(\hat{q}_1(\alpha_1); \hat{\theta}, \hat{\theta}) = C_1(\alpha_1; \hat{\theta}) = \hat{\alpha}_1^\dagger, \quad (4.1)$$

$$G_c^i(\hat{q}_i(\alpha_i) | z_{(i-1)}; \hat{\theta}, \hat{\theta}) = \frac{\partial_1^{i-1} C_i \{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1} | z_{(i-2)}; \hat{\theta}), \alpha_i; \hat{\theta}\}}{\partial_1^{i-1} C_{i-1} \{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1} | z_{(i-2)}; \hat{\theta}); \hat{\theta}\}} = \hat{\alpha}_i^\dagger, \quad (4.2)$$

for $i = 2, \dots, m$. Notice that $\hat{\alpha}_1^\dagger$ and $\hat{\alpha}_i^\dagger$, $i = 2, \dots, m$, correspond, respectively, to the plug-in estimators for the marginal coverage probability of $\hat{q}_1(\alpha_1)$ and for the conditional coverage probability of $\hat{q}_i(\alpha_i)$ given $Z_{(i-1)} = z_{(i-1)}$. The following proposition holds.

Proposition 4.1 *Under suitable regularity assumptions, assuring \sqrt{n} -consistency for the maximum likelihood estimator $\hat{\theta}$, the prediction limits*

$$\hat{q}_i^\dagger(\alpha_i) = 2\hat{q}_i(\alpha_i) - \hat{q}_i(\hat{\alpha}_i^\dagger), \quad i = 1, \dots, m, \quad (4.3)$$

with $\hat{q}_i(\alpha_i)$, $i = 1, \dots, m$, the estimative prediction limits which define (2.1) and $\hat{\alpha}_i^\dagger$, $i = 1, \dots, m$, given by (4.1) and (4.2), are such that

$$\hat{q}_i^\dagger(\alpha_i) = \hat{q}_i^c(\alpha_i) + o_p(n^{-1}), \quad i = 1, \dots, m,$$

where $\hat{q}_i^c(\alpha_i)$, $i = 1, \dots, m$, are specified as solutions to (3.5) with $\theta = \hat{\theta}$.

PROOF See Appendix B.

The modified estimative prediction limits $\hat{q}_i^\dagger(\alpha_i)$, $i = 1, \dots, m$, define suitable high-order asymptotic approximations for the plug-in calibrated prediction limits $\hat{q}_i^c(\alpha_i)$, $i = 1, \dots, m$, introduced in Section 3, and they are asymptotically equivalent to those ones specified by Corcuera and Giummolè (2006), achieving a coverage probability $\alpha + o(n^{-1})$. For $m = 1$, (4.3)

gives the modified estimative prediction limit introduced by Ueki and Fueda (2007) for a future one-dimensional random variable Z .

The advantage in using (4.3) is that we have a simple, high-order equivalent form for the calibrated prediction limits, which calculation does not require neither complicated asymptotic expansions, as in Corcuera and Giummolè (2006), nor the complete specification of the associated predictive distribution functions. In fact, to obtain the limits (4.3), we only need the computation of $\hat{\alpha}_i^\dagger$, $i = 1, \dots, m$. Whenever an explicit expression is not available, this can be done using the parametric bootstrap method; that is, we consider the bootstrap estimates given by

$$\hat{\alpha}_{1b}^\dagger = \hat{C}_1(\alpha_1; \theta)_b, \quad (4.4)$$

$$\hat{\alpha}_{ib}^\dagger = \frac{\partial_1^{i-1} \hat{C}_i \{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}), \alpha_i; \theta\}_b}{\partial_1^{i-1} \hat{C}_{i-1} \{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}); \theta\}_b}, \quad i = 2, \dots, m, \quad (4.5)$$

with $\hat{C}_1(\cdot)_b$, $\partial_1^{i-1} \hat{C}_i(\cdot)_b$ and $\partial_1^{i-1} \hat{C}_{i-1}(\cdot)_b$ specified in Section 3.

5 Dependent observations

A natural generalization of the procedure presented in the previous section concerns the case with dependent Y and Z . This includes the situation with $Y_1, \dots, Y_n, Z_1, \dots, Z_m$ continuous random variables, defined within a time series model. In this context, prediction is based on the conditional distribution of $Z = (Z_1, \dots, Z_m)$ given $Y = y$, with density and distribution functions $g(z|y; \theta)$ and $G(z|y; \theta)$, respectively, and the aim is to define an α -prediction region $R(Y, \alpha) \subset \mathbf{R}^m$ for Z such that

$$P_{Y,Z} \{Z \in R(Y, \alpha); \theta\} = E_Y [P_{Z|Y} \{Z \in R(Y, \alpha) | Y; \theta\}; \theta] = \alpha, \quad (5.1)$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0, 1)$, with $P_{Z|Y}\{\cdot|Y; \theta\}$ the conditional distribution of Z given Y .

We assume that there exists a transitive statistic (see Barndorff-Nielsen and Cox, 1996), namely a statistic $U = U(Y)$, with a fixed small dimension independent of the sample size n , such that Y and Z are conditional independent given U . This happens, for example, when we consider a time series model or a spatial model with a Markovian-type dependence structure. In this case, the conditional density of Z given $Y = y$ depends on y only via u , the observed value of U , so that $g(z|y; \theta) = g(z|u; \theta)$ and $G(z|y; \theta) = G(z|u; \theta)$. Consequently, it seems natural to consider a conditional version of (5.1) and thus to require that, for all θ , the conditional coverage probability equals the value α , that is

$$P_{Z,Y|U}\{Z \in R(Y, \alpha)|U = u; \theta\} = E_{Y|U}[P_{Z|U}\{Z \in R(Y, \alpha)|U; \theta\}|U = u; \theta] = \alpha,$$

for all $\alpha \in (0, 1)$, where the probability is calculated with respect to the conditional distribution of (Y, Z) given $U = u$ and the expectation is with respect to Y given $U = u$. Of course, conditional solutions satisfy also condition (5.1) and they are in some settings much easier to find.

Let us consider the estimative prediction region $R_e(Y, \alpha)$, based on the system of estimative prediction limits; in the presence of a transitive statistic U , the estimative prediction limit $\hat{q}_i(\alpha_i) = q_i(\alpha_i, z_{(i-1)}, u; \hat{\theta})$, $i = 1, \dots, m$, is the α_i -quantile of the conditional distribution of Z_i given $Z_{(i-1)} = z_{(i-1)}$ and $U = u$, evaluated at $\theta = \hat{\theta}$. The associated distribution and density functions are, respectively, $G^i = G^i(z_i|z_{(i-1)}, u; \theta)$ and $g^i = g^i(z_i|z_{(i-1)}, u; \theta)$. Note that, whenever $i = 1$, $\hat{q}_1(\alpha_1) = q_1(\alpha_1, u; \hat{\theta})$, $G^1 = G^1(z_1|u; \theta)$ and $g^1 = g^1(z_1|u; \theta)$. The conditional

coverage probability of $R_e(Y, \alpha)$ is

$$\begin{aligned} P_{Z,Y|U}\{Z \in R_e(Y, \alpha)|U = u; \theta\} &= E_{Y|U}[P_{Z|U}\{Z \in R_e(Y, \alpha)|U; \theta\}|U = u; \theta] \\ &= E_{Y|U}\left\{\int_{-\infty}^{\hat{q}_1(\alpha_1)} \cdots \int_{-\infty}^{\hat{q}_m(\alpha_m)} g(z|U; \theta) dz | U = u; \theta\right\} = C_m(\alpha_1, \dots, \alpha_m, u; \theta). \end{aligned}$$

Function $C_m(\alpha_1, \dots, \alpha_m, u; \theta)$ does not match the target value α and there is a coverage error term of order $O(n^{-1})$, as in the independence case. By substituting α_i with $G^i(z_i|z_{(i-1)}, u; \hat{\theta})$, $i = 1, \dots, m$, in $C_m(\alpha_1, \dots, \alpha_m, u; \theta)$, we obtain the following calibrated predictive distribution function

$$G_c(z|u; \hat{\theta}, \theta) = C_m\{G^1(z_1|u; \hat{\theta}), \dots, G^m(z_m|z_{(m-1)}, u; \hat{\theta}), u; \theta\}$$

and, after differentiation, the associated density function $g_c(z|u; \hat{\theta}, \theta)$. This predictive distribution gives, as quantiles of the corresponding conditional distribution functions, well-calibrated prediction limits. Thus, it is specified a prediction region which improves the estimative one, having unconditional and conditional coverage probability closer to the target nominal value α , for all $\alpha \in (0, 1)$. The proof is analogous to that one outlined for the independence case.

As in Section 3, since θ is unknown, we may use the corresponding plug-in estimators $G_c(z|u; \hat{\theta}, \hat{\theta})$ and $g_c(z|u; \hat{\theta}, \hat{\theta})$. If neither an explicit expression for the conditional coverage probability nor a first-order approximation is available, we may introduce a suitable parametric bootstrap estimator for $G_c(z|u; \hat{\theta}, \theta)$ and for the associate conditional distribution functions. The procedure is similar to that one considered in the case of independent observations, taking into account that the parametric bootstrap samples are now generated from the conditional distribution of Y given the observed value of the transitive statistic U , assuming $\theta = \hat{\theta}$. Under this respect, it could be useful to consider the simulation method proposed by Kabaila (1999) for estimating conditional expectations. Furthermore, the simplified procedure for computing

the improved prediction limits presented in Section 4 may be considered in this more general framework as well.

6 An application to a simple autoregressive model

The multivariate prediction regions considered in this paper are based on systems of conditional prediction limits, so that they present a peculiar non-rectangular form. This specification is not frequent in the time series framework, where prediction region are usually of rectangular form in order to derive simple marginal statements about the single components of the future random vector Z . However, the current prediction regions turn out to be extremely useful for forecasting future paths of time series observations as well, since they provide flexible predictive statements, which may account for alternative path forecast scenarios. In this section we present an application concerning autoregressive time series models and we shall perform a simple simulation study in order to emphasize the superiority of the improved solution over the estimative one.

Let us consider a stationary, first-order autoregressive (AR) process $\{Y_k\}_{k \geq 1}$ defined as

$$Y_k = \mu + \rho(Y_{k-1} - \mu) + \varepsilon_k, \quad k \geq 1,$$

where $\mu \in \mathbf{R}$, $|\rho| < 1$ and $\{\varepsilon_k\}_{k \geq 1}$ is a sequence of independent normal distributed random variables with zero mean and variance $\sigma^2 > 0$. Adopting the usual notation, $Y = (Y_1, \dots, Y_n)$, $Z = (Z_1, \dots, Z_m) = (Y_{n+1}, \dots, Y_{n+m})$ and $\theta = (\theta_1, \theta_2, \theta_3) = (\mu, \rho, \sigma^2)$ is the unknown parameter. In this case, the observable random vector Y and the future random vector Z are dependent and the transitive statistic is $U = Y_n$. Indeed, likelihood inference is conditioned on $Y_0 = y_0$,

with y_0 known, and the maximum likelihood estimators are given explicitly and correspond to

$$\hat{\mu} = \frac{\sum_{i=1}^n Y_i - \hat{\rho} \sum_{i=1}^n Y_{i-1}}{n(1-\hat{\rho})}, \quad \hat{\rho} = \frac{\sum_{i=1}^n Y_i Y_{i-1} - n^{-1} \sum_{i=1}^n Y_i \sum_{i=1}^n Y_{i-1}}{\sum_{i=1}^n Y_{i-1}^2 - n^{-1} (\sum_{i=1}^n Y_{i-1})^2}, \quad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (Y_i - \hat{\mu}_i)^2,$$

with $\hat{\mu}_i = \hat{\mu} + \hat{\rho}(Y_{i-1} - \hat{\mu})$, $i = 1, \dots, n$.

Moreover, it is easy to see that Z_i given $Z_{i-1} = z_{i-1}$, $i = 1, \dots, m$, follows a normal distribution with mean $\mu_{n+i} = \mu + \rho(z_{i-1} - \mu)$ and variance σ^2 . Thus, $g^i(z_i|z_{i-1}; \theta) = \sigma^{-1} \phi\{(z_i - \mu_{n+i})/\sigma\}$ and $G^i(z_i|z_{i-1}; \theta) = \Phi\{(z_i - \mu_{n+i})/\sigma\}$, $i = 1, \dots, m$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the density and the distribution function of a standard normal random variable. Note that for $i = 1$ the conditional event corresponds to $(Z_0 = z_0) = (Y_n = y_n)$.

For this simple stationary time series model, the estimative prediction region $R_e(Y, \alpha)$ is defined by (2.1), with system of conditional prediction limits given by $\hat{q}_i(\alpha_i) = q_i(\alpha_i, z_{i-1}, \hat{\theta}) = \hat{\mu}_{n+i} + u_{\alpha_i} \hat{\sigma}$, $i = 1, \dots, m$, with $\hat{\mu}_{n+i} = \hat{\mu} + \hat{\rho}(z_{i-1} - \hat{\mu})$ and u_{α_i} such that $\Phi(u_{\alpha_i}) = \alpha_i$, where $\prod_{i=1}^m \alpha_i = \alpha$. In order to improve the coverage accuracy of $R_e(Y, \alpha)$ we may consider the modified prediction limits associated to the improved multivariate predictive density introduced by Corcuera and Giummolè (2006). This solution, although in this simple model is explicitly available, is usually based on complicated asymptotic calculations and often unfeasible.

For this AR(1) model, we consider the alternative easier solution introduced in Section 4, extended to the case of dependent observations, and obtained by generalizing the Ueki and Fueda's procedure to multivariate prediction. For evaluating the quantities $\hat{\alpha}_i^\dagger$, $i = 1, \dots, m$, we use the parametric bootstrap estimators, specified by (4.4) and (4.5), based on 2,000 bootstrap samples generated keeping fixed the observed value of the transitive statistic Y_n .

The following simple simulation study compares the performance of this simplified improved predictive solution with that one of the estimative solution. Conditional coverage probabilities

for the estimative and the improved conditional prediction limits of level $\alpha = 0.9, 0.95$ are calculated by means of the conditional simulation technique presented in Kabaila (1999), keeping fixed the last observed value y_n . The estimates of the conditional coverage probabilities are based on 2,000 samples of dimension $n = 50, 100$ simulated from an AR(1) model with the last observation fixed to $y_n = 1$ and assuming $y_0 = 0$; indeed, we consider $\mu = 1$, $\sigma^2 = 1$ and (a) $\rho = 0.5$, (b) $\rho = 0.75$. The prediction regions have dimension $m = 5, 10, 25, 50$ and $\alpha_i = \alpha^{1/m}$, $i = 1, \dots, m$. Similar results are obtained with alternative values for the observations y_n, y_0 and for the model parameters. The results are collected in Table 1 and show that the improved procedure remarkably improves on the estimative one. The improvement is more pronounced when the dimension m of the future random vector is high with respect to n , since the performance of the estimative solution progressively declines. On the other hand, the improved solution maintains a good coverage accuracy also in the challenging situations where $m = 50$ for $n = 50$ and $n = 100$. Thus, in accordance with the theoretical findings, the improved prediction limits defined in Section 4 can be fruitfully considered for making multivariate prediction statements, as a valid and simpler alternative to the improved methods based on asymptotic analytic calculations.

Table 1 here

A Appendix

We present the outline of the proof of Proposition 3.1.

By considering the explicit expression for the $O(n^{-1})$ coverage error term of the estimative prediction region $R_e(Y, \alpha)$, determined by Corcuera and Giummolè (2006, Corollary 2), we

state that

$$C_m(\alpha_1, \dots, \alpha_m; \theta) = \alpha + Q_m(\alpha_1, \dots, \alpha_m; \theta) + o(n^{-1}), \quad (\text{A.1})$$

with $\alpha = \prod_{i=1}^m \alpha_i$ and the $O(n^{-1})$ term

$$\begin{aligned} Q_m(\alpha_1, \dots, \alpha_m; \theta) &= - \sum_{i=1}^m b_r(\theta) \left\{ \int_{-\infty}^{q_1} \cdots \int_{-\infty}^{q_{i-1}} G^i(q_i | z_{(i-1)}; \theta)_r \right\} \prod_{h=i+1}^m \alpha_h \\ &+ \frac{1}{2} \sum_{i=1}^m i^{rs}(\theta) \left[\int_{-\infty}^{q_1} \cdots \int_{-\infty}^{q_{i-1}} \left\{ [2] \frac{g^i(q_i | z_{(i-1)}; \theta)_r G^i(q_i | z_{(i-1)}; \theta)_s}{g^i(q_i | z_{(i-1)}; \theta)} - G^i(q_i | z_{(i-1)}; \theta)_{rs} \right\} \right] \prod_{h=i+1}^m \alpha_h \\ &+ \frac{1}{2} \sum_{i=1}^m \sum_{j < i} i^{rs}(\theta) [2] \left\{ \int_{-\infty}^{q_1} \cdots \int_{-\infty}^{q_{j-1}} \int_{-\infty}^{q_{j+1}(j)} \cdots \int_{-\infty}^{q_{i-1}(j)} G^j(q_j | z_{(j-1)}; \theta)_r G^i(q_i | z_{(i-1)}(j); \theta)_s \right\} \prod_{h=i+1}^m \alpha_h, \end{aligned}$$

where we put $\prod_{h=i+1}^m \alpha_h = 1$, whenever $i = m$. Here,

$$\int_{-\infty}^{q_k} = \int_{-\infty}^{q_k} g^k(z_k | z_{(k-1)}; \theta) dz_k, \quad \int_{-\infty}^{q_k(j)} = \int_{-\infty}^{q_k(j)} g^k(z_k | z_{(k-1)}(j); \theta) dz_k,$$

$q_k = q_k(\alpha_k, z_{(k-1)}; \theta)$ and $q_k(j) = q_k(\alpha_k, z_{(k-1)}(j); \theta)$ is the α_k -quantile of $G^k(z_k | z_{(k-1)}(j); \theta)$, with $z_{(k-1)}(j) = (z_1, \dots, z_{j-1}, q_j, z_{j+1}, \dots, z_{k-1})$. Indeed, $G^i(\cdot | \cdot; \theta)_r$ and $G^i(\cdot | \cdot; \theta)_{rs}$, are the first and the second partial derivatives of $G^i(\cdot | \cdot; \theta)$ with respect to the corresponding components of vector θ . Partial differentiation of both sides of (A.1) with respect to $\alpha_1, \dots, \alpha_m$ gives

$$\partial_1^m C_m(\alpha_1, \dots, \alpha_m; \theta) = 1 + \partial_1^m Q_m(\alpha_1, \dots, \alpha_m; \theta) + o(n^{-1})$$

and the consequent substitution in formula (3.2), evaluated at $\theta = \hat{\theta}$, with $\partial_1^m Q_m(\alpha_1, \dots, \alpha_m; \hat{\theta})$ written explicitly, completes the proof.

B Appendix

We present the proof of Proposition 4.1.

From (A.1), it is easy to see that

$$\hat{\alpha}_1^\dagger = C_1(\alpha_1; \hat{\theta}) = \alpha_1 + Q_1(\alpha_1; \hat{\theta}) + o_p(n^{-1}),$$

with

$$Q_1(\alpha_1; \hat{\theta}) = -b_r(\hat{\theta})G^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}_r + \frac{1}{2}i^{rs}(\hat{\theta}) \left\{ [2] \frac{g^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}_r G^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}_s}{g^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}} - G^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}_{rs} \right\}.$$

Indeed, after some algebra, we obtain that, for $i = 2, \dots, m$,

$$\begin{aligned} \hat{\alpha}_i^\dagger &= \frac{\partial_1^{i-1} C_i\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}), \alpha_i; \hat{\theta}\}}{\partial_1^{i-1} C_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}); \hat{\theta}\}} \\ &= \frac{\alpha_i + \partial_1^{i-1} Q_i\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}), \alpha_i; \hat{\theta}\} + o_p(n^{-1})}{1 + \partial_1^{i-1} Q_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}); \hat{\theta}\} + o_p(n^{-1})} \\ &= \alpha_i + \partial_1^{i-1} Q_i\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}), \alpha_i; \hat{\theta}\} \\ &\quad - \alpha_i \partial_1^{i-1} Q_{i-1}\{G^1(z_1; \hat{\theta}), \dots, G^{i-1}(z_{i-1}|z_{(i-2)}; \hat{\theta}); \hat{\theta}\} + o_p(n^{-1}) \\ &= \alpha_i + Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta}) + o_p(n^{-1}), \end{aligned}$$

with

$$\begin{aligned} Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta}) &= -b_r(\hat{\theta})G^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}_r \\ &\quad + \frac{1}{2}i^{rs}(\hat{\theta}) \left[[2] \frac{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}_r G^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}_s}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} - G^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}_{rs} \right] \\ &\quad + \frac{1}{2} \sum_{j < i} i^{rs}(\hat{\theta}) [2] \frac{d/dz_j [G^j\{z_j|z_{(j-1)}; \hat{\theta}\} \prod_{k=j+1}^{i-1} g^k\{z_k|z_{(k-1)}; \hat{\theta}\} G^i\{z_i|z_{(i-1)}; \hat{\theta}\}]}{\prod_{k=j}^{i-1} g^k\{z_k|z_{(k-1)}; \hat{\theta}\}}. \end{aligned}$$

Using a stochastic Taylor expansion for $\hat{q}_i(\hat{\alpha}_i^\dagger)$ around $\hat{\alpha}_i^\dagger = \alpha_i$, we obtain that

$$\begin{aligned} \hat{q}_i(\hat{\alpha}_i^\dagger) &= \hat{q}_i(\alpha_i) + \frac{\hat{\alpha}_i^\dagger - \alpha_i}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} + o_p(n^{-1}) \\ &= \hat{q}_i(\alpha_i) + \frac{Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta})}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} + o_p(n^{-1}), \quad i = 1, \dots, m. \end{aligned}$$

Hereafter, for $i = 1$, $Q_1(\alpha_1; \hat{\theta})$ is considered instead of $Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta})$ and $g^1\{\hat{q}_1(\alpha_1); \hat{\theta}\}$ instead of $g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}$. Substitution in (4.3) gives

$$\begin{aligned}\hat{q}_i^\dagger(\alpha_i) &= 2\hat{q}_i(\alpha_i) - \hat{q}_i(\hat{\alpha}_i^\dagger) = \hat{q}_i(\alpha_i) + \hat{q}_i(\alpha_i) - \hat{q}_i(\hat{\alpha}_i^\dagger) \\ &= \hat{q}_i(\alpha_i) - \frac{Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta})}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} + o_p(n^{-1}), \quad i = 1, \dots, m,\end{aligned}\tag{B.2}$$

which corresponds, up to terms of order $O_p(n^{-1})$, to the modified estimative prediction limit introduced by Corcuera and Giummolè (2006, Corollary 3). Furthermore, since $\hat{q}_i^c(\alpha_i)$, $i = 1, \dots, m$, is such that $G_c^i\{\hat{q}_i^c(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\} = \alpha_i$, with a suitable stochastic Taylor expansion around $\hat{q}_i^c(\alpha_i) = \hat{q}_i(\alpha_i)$, we have that

$$\begin{aligned}\alpha_i &= G_c^i\{\hat{q}_i^c(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\} = G_c^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\} \\ &\quad + \{\hat{q}_i^c(\alpha_i) - \hat{q}_i(\alpha_i)\}g_c^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\} + o_p(n^{-1}).\end{aligned}$$

Finally, recalling that $G_c^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\} = \hat{\alpha}_i^\dagger$ and that $g_c^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}, \hat{\theta}\}$ is first-order equivalent to $g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}$, we obtain that

$$\begin{aligned}\hat{q}_i^c(\alpha_i) &= \hat{q}_i(\alpha_i) - \frac{\hat{\alpha}_i^\dagger - \alpha_i}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} + o_p(n^{-1}) \\ &= \hat{q}_i(\alpha_i) - \frac{Q_{i|(i-1)}(\alpha_i, z_{(i-1)}; \hat{\theta})}{g^i\{\hat{q}_i(\alpha_i)|z_{(i-1)}; \hat{\theta}\}} + o_p(n^{-1}), \quad i = 1, \dots, m.\end{aligned}\tag{B.3}$$

Therefore, from (B.2) and (B.3), we may conclude that $\hat{q}_i^\dagger(\alpha_i) = \hat{q}_i^c(\alpha_i) + o_p(n^{-1})$, $i = 1, \dots, m$.

References

- [1] Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

- [2] Barndorff-Nielsen, O.E. and Cox, D.R. (1996). Prediction and asymptotics. *Bernoulli* **2**, 319-340.
- [3] Beran, R. (1990). Calibrating prediction regions. *J. Amer. Statist. Assoc.* **85**, 715-723.
- [4] Corcuera, J.M. and Giummolè, F. (2006). Multivariate prediction. *Bernoulli* **12**, 157-168.
- [5] Fonseca, G., Giummolè, F. and Vidoni, P. (2014). Calibrating predictive distributions. *J. Stat. Comput. Simul.* **84**, 373-383.
- [6] Kabaila, P. (1999). An efficient simulation method for the computation of a class of conditional expectations. *Aust. N. Z. J. Stat.* **41**, 331-336.
- [7] Hall, P., Peng, L. and Tajvidi, N. (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika* **86**, 871-880.
- [8] Lawless, J.F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92**, 529-542.
- [9] Ueki, M. and Fueda, K. (2007). Adjusting estimative prediction limits. *Biometrika* **94**, 509-511.
- [10] Vidoni, P. (1998). A note on modified estimative prediction limits and distributions. *Biometrika* **85**, 949-953.
- [11] Vidoni, P. (2009). Improved prediction intervals and distribution functions. *Scand. J. Stat.* **36**, 735-748.

			(a)		(b)	
α	n	m	Estimative	Improved	Estimative	Improved
0.9	50	5	0.862	0.896	0.858	0.901
		10	0.837	0.890	0.836	0.897
		25	0.812	0.890	0.806	0.879
		50	0.794	0.888	0.789	0.889
	100	5	0.886	0.907	0.876	0.896
		10	0.868	0.892	0.874	0.904
		25	0.872	0.906	0.866	0.898
		50	0.850	0.896	0.836	0.894
0.95	50	5	0.912	0.944	0.919	0.944
		10	0.911	0.945	0.902	0.943
		25	0.890	0.948	0.879	0.945
		50	0.887	0.949	0.861	0.933
	100	5	0.930	0.942	0.927	0.942
		10	0.932	0.951	0.923	0.940
		25	0.924	0.948	0.926	0.949
		50	0.911	0.940	0.925	0.958

Table 1: AR(1) Gaussian model with $\mu = 1$, $\sigma^2 = 1$ and (a) $\rho = 0.5$, (b) $\rho = 0.75$. Conditional coverage probabilities for estimative and improved, bootstrap-based, prediction limits of level $\alpha = 0.9, 0.95$, with $m = 5, 10, 25, 50$. Estimation is based on 2,000 Monte Carlo conditional (on $y_n = 1$) samples of dimension $n = 50, 100$, with $y_0 = 0$. Bootstrap procedure is based on 2,000 conditional bootstrap samples. Estimated standard errors are always smaller than 0.0092.