Università degli studi di Udine

Improved multivariate prediction regions for Markov process models

Original

Availability:
This version is available http://hdl.handle.net/11390/1100394 since 2020-03-21T19:01:39Z

Publisher:

Published
DOI:10.1007/s10260-016-0362-y

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)

25 April 2024

Improved multivariate prediction regions for Markov process models

Paolo Vidoni

Department of Economics and Statistics, University of Udine

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

paolo.vidoni@uniud.it

Abstract

This paper concerns the specification of multivariate prediction regions which may be useful in time series applications whenever we aim at considering not just one single forecast but a group of consecutive forecasts. We review a general result on improved multivariate prediction and we use it in order to calculate conditional prediction intervals for Markov process models so that the associated coverage probability turns out to be close to the target value. This improved solution is asymptotically superior to the estimative one, which is simpler but it may lead to unreliable predictive conclusions. An application to general autoregressive models is presented, focusing in particular on AR and ARCH models.

Keywords: autoregressive model; coverage probability; estimative prediction limits; simultaneous prediction limits; time series.

1 Introduction

Predictive inference concerning a future multivariate random variable, with particular regard to the specification of multivariate prediction regions, may be of considerable interest in time series applications whenever we aim at considering not just one single forecast but a group of consecutive forecasts. In such cases we usually look for a prediction region, obtained from a system of prediction intervals, which covers the future observations of the process with an overall specified coverage probability α. The purpose of this paper is to employ the general results given by Corcuera and Giummolè (2006) in order to calculate prediction regions for Markov process models so that the associated coverage probability turns out to be close to the target value α. The proposed prediction regions are analytically defined and they are not necessarily of rectangular form, since each component prediction interval depends on the previous future potential observations. Even though the non-rectangular form does not automatically produce marginal predictive statements for the single future observations, it enables the specification of a system of conditional prediction intervals accounting for alternative path forecast scenarios for the future realizations of the process.

Let us consider a one-dimensional, discrete-time, stochastic process $\{Y_n\}_{n\geq 1}$ and let us suppose that the random vector $Y = (Y_1, \ldots, Y_n)$, $n \geq 1$, is observed. We aim at predicting a future, or yet unobserved, random vector $Z = (Z_1, \ldots, Z_m) = (Y_{n+1}, \ldots, Y_{n+m})$, $m \geq 1$; (Y, Z) is a continuous random vector following a joint density $p(y, z; \theta)$, with $\theta \in \Theta \subseteq \mathbf{R}^d$ an unknown d-dimensional parameter. We require the existence of a transitive statistic (Barndorff-Nielsen and Cox 1996; Bjørnstad 1996) $U = U(Y)$, with a fixed dimension, independent of the sample size n, so that Y and Z are conditionally independent given U; in the case of a Markov

process of order $k \geq 1$, we have $U = (Y_{n-k+1}, \ldots, Y_n)$. Thus, the conditional distribution of Z given $Y = y$ depends on y only via u, the observed value of U; the conditional density of Z given $Y = y$ is $g(z|y;\theta) = g(z|u;\theta)$, and analogously for the conditional distribution function $G(z|y;\theta) = G(z|u;\theta)$. We also assume that $p(y,z;\theta)$, $g(z|y;\theta)$ and $G(z|y;\theta)$ are sufficiently smooth functions of the parameter θ, for fixed y and z.

Although prediction problems may be considered from alternative perspectives, in this paper we follow the frequentist approach and the objective is to define a suitable α-prediction region $R(Y,\alpha) \subset \mathbf{R}^m$ for Z, such that

$$P_{Y,Z|U}\{Z \in R(Y,\alpha)|U = u;\theta\} = \alpha, \tag{1}$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0,1)$. The above probability is called conditional coverage probability and it is calculated with respect to the conditional joint distribution of (Z,Y) given $U = u$. It is immediate to note that, if $R(Y,\alpha)$ satisfies (1) then it has also a marginal coverage probability equal to the target value α, that is $P_{Y,Z}\{Z \in R(Y,\alpha);\theta\} = \alpha$, for all θ. The use of the coverage probability conditional on $U = u$, as a suitable measure of confidence of prediction methods, involves hypothetical or real replications of the sample experiment considering U as fixed to the observed value u. For a detailed and convincing justification of the conditional approach see Kabaila (1999a) and Vidoni (2004).

The simplest way for making prediction on Z is by using the estimative predictive density $g(z|u;\widehat{\theta})$, where the unknown parameter θ is substituted with an asymptotically efficient estimator $\widehat{\theta}$ based on Y, such that $\widehat{\theta} - \theta = O_p(n^{-1/2})$; we usually consider the maximum likelihood estimator or any asymptotically equivalent alternative estimator. However, prediction regions based on the estimative procedure are not entirely adequate predictive solutions, since they

underestimate the additional uncertainty introduced by assuming $\theta = \widehat{\theta}$. Thus, 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 and/or large m.

Concerning the univariate case, Barndorff-Nielsen and Cox (1996), Ueki and Fueda (2007) and Vidoni (1998) 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 $o(n^{-1})$. Alternative approaches, based on a suitable calibrating procedure (Beran 1990; Hall et al. 1999) or on the specification of an approximate pivotal quantity (Lawless and Fredette 2005) have been also proposed for improving the estimative solution. An application of these solutions giving improved prediction intervals for time series models, with particular regard to autoregressive processes, can be found in Corcuera (2008), Kabaila and He (2004), Kabaila and Syuhada (2007) and Vidoni (2004, 2009).

With regard to the multivariate case, the calibrating approach introduced by Beran (1990), and applied for example by Hall et al. (1999) using a bootstrap simulation procedure, improves the estimative prediction regions as well. This is a very general approach and it consists in finding a suitable value $\bar{\alpha}$ such that the coverage probability of the estimative, recalibrated, $\bar{\alpha}$-prediction region is equal or close to the target value α. With particular attention to the time series framework, Jordà and Marcellino (2010) propose a rectangular prediction region, suitably corrected in order to control the coverage probability. This prediction region is obtained from the Scheffé's (1959) one, which has originally an elliptical form. Wang and Fa (2010) consider multivariate linear mixed models for the analysis of longitudinal data and specify simultaneous predictions intervals, simply obtained by using the estimative procedure. In a

similar context, Concordet and Servien (2014) introduce suitable corrections with the aim of improving the coverage accuracy of the estimative solution. Recently, Wolf and Wunderly (2015) consider a general, simulation-based approach for the construction of joint prediction regions that will contain the entire future path with the desired coverage probability. Their procedure is essentially bootstrap-based, it is rather general, it does not necessarily require the specification of the distribution of the error terms and it specifies joint rectangular prediction regions.

In a more general context, encompassing the time series framework, Corcuera and Giummolè (2006) introduce a procedure, based on asymptotic calculations, in order to define a system of improved prediction intervals with an overall coverage error of smaller asymptotic order than that one of the estimative solution; they also specify the associated multivariate predictive density. In the present paper, we follow the Corcuera and Giummolè's approach and we apply their results for the specification of improved multivariate prediction regions within Markov process models. Differently from the results previously reviewed, the proposed prediction regions have a non-rectangular form. We present also an application to general autoregressive time series models, focusing in particular on autoregressive (AR) models and on autoregressive conditional heteroscedastic (ARCH) models, and we perform a simple simulation study in order to emphasize the improvement achieved by our solution, in terms of coverage probability, over the estimative one.

2 Review on improved univariate prediction limits for Markov processes

Let us review the Barndorff-Nielsen and Cox's (1996) asymptotic procedure for improving estimative prediction intervals, as specified for the situation with dependent observations by Vidoni (2004). Let us consider the univariate case, so that $m = 1$, and let us assume that $U = (Y_{n-k+1}, \ldots, Y_n)$, $k \geq 1$. The focus is, in particular, on the estimative α-prediction limit $z_\alpha(\widehat{\theta}, u)$ for $Z = Y_{n+1}$, defined as the α-quantile of the estimative conditional distribution function $G(z|u;\widehat{\theta})$. It is well known that the conditional and the unconditional coverage probability of $z_\alpha(\widehat{\theta}, u)$ differs from α by an error term of order $O(n^{-1})$, which may be substantial, leading to unreliable predictive conclusions. A general solution for improving estimative prediction limits involves higher-order asymptotic corrections. The key idea is to obtain a suitable modification of $z_\alpha(\widehat{\theta}, u)$, defined in order to absorb the additional $O(n^{-1})$ term in the associated coverage probability, so that the error term turns out to be reduced to order $O(n^{-3/2})$.

Let us use the Einstein summation convention, so that if an index occurs more than once in a summand then summation over that index is understood, and let θ_r and $\widehat{\theta}_r$, $r = 1, \ldots, d$, denote the r-th component of θ and $\widehat{\theta}$, respectively. Since U is a transitive statistic, the conditional coverage probability of the estimative prediction limit may be rewritten as

$$P_{Y,Z|U}\{Z \leq z_\alpha(\widehat{\theta}, U)|U = u; \theta\} = E_{Y|U}[G\{z_\alpha(\widehat{\theta}, U)|U; \theta\}|U = u; \theta],$$

where the expectation is with respect to the conditional distribution of Y given $U = u$. Using a stochastic Taylor expansion in $\widehat{\theta} = \theta$, we have that

$$P_{Y,Z|U}\{Z \leq z_\alpha(\widehat{\theta}, U)|U = u; \theta\} = \alpha + Q\{z_\alpha(\theta, u), u; \theta\} + O(n^{-3/2}), \tag{2}$$

with

$$Q(z,u;\theta) = -b_r(\theta,u)G_r(z|u;\theta) - \frac{1}{2}i^{rs}(\theta,u)\left\{G_{rs}(z|u;\theta) + [2]G_r(z|u;\theta)\ell_s(\theta;u,z)\right\},$$

where [2] indicates the sum of two terms obtained by permutation of indices r and s. Here, $b_r(\theta,u)$ and $i^{rs}(\theta,u)$ are such that

$$\begin{aligned} E_{Y|U}\{(\widehat{\theta}-\theta)_r|U=u;\theta\} &= b_r(\theta,u) + O(n^{-3/2}), \\ E_{Y|U}\{(\widehat{\theta}-\theta)_r(\widehat{\theta}-\theta)_s|U=u;\theta\} &= i^{rs}(\theta,u) + O(n^{-3/2}), \end{aligned}$$

while $G_r(z|u;\theta)$ and $G_{rs}(z|u;\theta)$ are the first and the second partial derivatives of $G(z|u;\theta)$, with respect to components θ_r and θ_s of θ, and $\ell_s(\theta;u,z) = \partial \log g(z|u;\theta)/\partial\theta_s$. For the $O(n^{-1})$ conditional bias term of $\widehat{\theta}_r$ holds the useful formula (Vidoni 2004)

$$b_r(\theta,u) = b_r(\theta) + \log p(u;\theta)_s i^{rs}(\theta) + o(n^{-1}), \tag{3}$$

where $b_r(\theta) = E_Y\{(\widehat{\theta}-\theta)_r;\theta\}$ is the $O(n^{-1})$ unconditional bias term of $\widehat{\theta}_r$ and $\log p(u;\theta)_s = \partial \log p(u;\theta)/\partial\theta_s$, with $p(u;\theta)$ the marginal density of U. Moreover, $i^{rs}(\theta,u) = i^{rs}(\theta) + O(n^{-3/2})$, with $i^{rs}(\theta)$ the (r,s)-element of the inverse of the expected information matrix based on Y.

A modification of the estimative prediction limit, with conditional (and unconditional) coverage error of order $O(n^{-3/2})$, is

$$\tilde{z}_\alpha(y) = z_\alpha(\widehat{\theta},u) - \frac{Q\{z_\alpha(\widehat{\theta},u),u;\widehat{\theta}\}}{g\{z_\alpha(\widehat{\theta},u)|u;\widehat{\theta}\}}. \tag{4}$$

Moreover, the predictive distribution function and the predictive density function which give (4) as α-quantile up to terms of order $O(n^{-1})$ are, respectively,

$$\tilde{G}(z;y) = G(z|u;\widehat{\theta}) + Q(z,u;\widehat{\theta}),$$

$$
\begin{aligned}
\tilde{g}(z;y) &= g(z|u;\widehat{\theta})\left\{1+\frac{dQ(z,u;\widehat{\theta})/dz}{g(z|u;\widehat{\theta})}\right\} \\
&= g\{z+R(z,u;\widehat{\theta})|u;\widehat{\theta}\}\left\{1+dR(z,u;\widehat{\theta})/dz\right\},
\end{aligned}
$$

with $R(z,u;\widehat{\theta}) = Q(z,u;\widehat{\theta})/g(z|u;\widehat{\theta})$.

Equation (4) enables the computation of univariate improved prediction intervals for a number of Markov process models, such as AR and ARCH models (see, for example, Corcuera 2008; Vidoni 2004). Note that, whenever the quantities $b_r(\theta,u)$ and $i^{rs}(\theta)$, $r,s=1,\ldots,d$, are not known explicitly, and feasible asymptotic approximations are not available, they can be computed by means of suitable (conditional) parametric bootstrap simulation procedures.

3 Improved prediction regions for Markov processes

Let us suppose, in order to simplify the exposition, that $\{Y_n\}_{n\geq 1}$ is a first-order Markov process. At the end of the present section we shall extend the results to the general case of a Markov process of order $k \geq 1$. Therefore, in this particular situation, Y and Z are conditionally independent given $U = Y_n$ and the conditional density of Z given $Y = y$ is

$$
g(z|y;\theta) = g(z|y_n;\theta) = \prod_{i=1}^{m} g(z_i|z_{i-1};\theta),
$$

where, with a slight abuse of notation, $g(z_i|z_{i-1};\theta)$ indicates the conditional density of Z_i given $Z_{i-1} = z_{i-1}$, $i=1,\ldots,m$, where $Z_0 = Y_n$ and $z_0 = y_n$. The associated conditional distribution functions are, respectively, $G(z|y;\theta) = G(z|y_n;\theta)$ and $G(z_i|z_{i-1};\theta)$, $i=1,\ldots,m$.

Let us consider the multivariate setting by assuming $m > 1$ fixed. A particular procedure for specifying prediction regions involves a system of simultaneous inequalities based on suitable

prediction limits, that is quantities $c_1(y)$, $c_i(y, z_1^{i-1})$, $i = 2, \ldots, m$, such that

$$P_{Y,Z|Y_n}\{Z_1 \leq c_1(Y), Z_2 \leq c_2(Y, Z_1^1), \cdots, Z_m \leq c_m(Y, Z_1^{m-1})|Y_n = y_n; \theta\} = \alpha, \tag{5}$$

for all θ, with $\alpha \in (0, 1)$; hereafter, $Z_h^v = (Z_h, \ldots, Z_v)$ and $z_h^v = (z_h, \ldots, z_v)$, $0 \leq h \leq v$. Note that the prediction limits $c_i(y, z_1^{i-1})$, $i = 2, \ldots, m$, depend on the past observations z_1^{i-1}. This sequential-type procedure for specifying a multivariate prediction region is especially useful for time series analysis, where there is a prescribed temporal ordering in the observed data. Moreover, this particular choice for the prediction limits determines a shape for the prediction region which is not necessarily rectangular, for the two-sided case, or semi-infinite rectangular, for the one-sided case.

Since θ is unknown, an intuitive approximate solution to (5) is given by the system of simultaneous estimative prediction limits $z_\alpha^i(\widehat{\theta}, z_{i-1})$, $i = 1, \ldots, m$, obtained by replacing θ with $\widehat{\theta}$ in $z_\alpha^i(\theta, z_{i-1})$, namely, the α_i-quantile of the conditional distribution of Z_i given $Z_{i-1} = z_{i-1}$, specified as the solution, with respect to z_i, of $G\{z_i|z_{i-1}; \theta\} = \alpha_i$. Indeed, the conditional coverage probabilities α_i, $i = 1, \ldots, m$, are such that $\prod_{i=1}^m \alpha_i = \alpha$. It is well-known that the estimative solution could be not adequate for prediction, being usually rather imprecise. This drawback maintains in the multivariate framework, since, as proved by Corcuera and Giummolè (2006), estimative prediction regions present a coverage error of order $O(n^{-1})$. Moreover, the procedure for correcting each univariate estimative prediction limit $z_\alpha^i(\widehat{\theta}, z_{i-1})$, $i = 1, \ldots, m$, outlined in the previous section with regard to the univariate case, fails to improve the overall coverage probability of the associated prediction region. For multivariate prediction, we have to take into account also the potential dependence between the components of vector Z and the interaction among the simultaneous estimative prediction limits, induced by the fact that

they are all evaluated assuming $\theta = \widehat{\theta}$.

Following Corcuera and Giummolè (2006), the simultaneous prediction limits which improve the estimative ones, since they satisfy (5) with an error term of order $O(n^{-3/2})$, correspond to

$$\tilde{z}^1_\alpha = \tilde{z}^1_\alpha(y) = z^1_\alpha(\widehat{\theta}, y_n) - \frac{h_1\{z^1_\alpha(\widehat{\theta}, y_n), y_n, z_0; \widehat{\theta}\}}{g\{z^1_\alpha(\widehat{\theta}, y_n)|y_n; \widehat{\theta}\}} \tag{6}$$

and, for $i = 2, \ldots, m$,

$$\tilde{z}^i_\alpha = \tilde{z}^i_\alpha(y, z^{i-1}_1) = z^i_\alpha(\widehat{\theta}, z_{i-1}) - \frac{h_i\{z^i_\alpha(\widehat{\theta}, z_{i-1}), y_n, z_{i-1}; \widehat{\theta}\}}{g\{z^i_\alpha(\widehat{\theta}, z_{i-1})|z_{i-1}; \widehat{\theta}\}} - \frac{\sum_{j=1}^{i-1} h_{ij}\{z^i_\alpha(\widehat{\theta}, z_{i-1}), z^{j+1}_{j-1}; \widehat{\theta}\}}{g\{z^i_\alpha(\widehat{\theta}, z_{i-1})|z_{i-1}; \widehat{\theta}\}}. \tag{7}$$

The term $h_i(z_i, y_n, z_{i-1}; \theta)$, $i = 1, \ldots, m$, is of order $O(n^{-1})$ and it is given by

$$\begin{aligned} h_i(z_i, y_n, z_{i-1}; \theta) = & - b_r(\theta, y_n) G_r(z_i|z_{i-1}; \theta) \\ & - \frac{1}{2} i^{rs}(\theta) \left\{ G_{rs}(z_i|z_{i-1}; \theta) - [2] G_r(z_i|z_{i-1}; \theta) \ell_s(\theta; z_{i-1}, z_i) \right\}, \end{aligned}$$

while the $O(n^{-1})$ term $h_{ij}(z_i, z^{j+1}_{j-1}; \theta)$ is given by

$$h_{ij}(z_i, z^{j+1}_{j-1}; \theta) = \frac{1}{2}[2] i^{rs}(\theta) G_r(z_i|z_{i-1}; \theta) \left\{ \ell_s(\theta; z_{j-1}, z_j) + \frac{G_s(z_j|z_{j-1}; \theta)\, d \log g(z_{j+1}|z_j; \theta)/dz_j}{g(z_j|z_{j-1}; \theta)} \right\},$$

for $j = 1, \ldots, i-2$, $i = 3, \ldots, m$, and by

$$h_{ij}(z_i, z^{j+1}_{j-1}; \theta) = \frac{1}{2}[2] i^{rs}(\theta) \left\{ G_r(z_i|z_{i-1}; \theta) \ell_s(\theta; z_j, z_{j+1}) + \frac{G_s(z_j|z_{j-1}; \theta)\, dG_r(z_i|z_{i-1}; \theta)/dz_{i-1}}{g(z_{i-1}|z_{i-2}; \theta)} \right\},$$

for $j = i-1$, $i = 2, \ldots, m$, where $b_r(\theta, y_n)$, $i^{rs}(\theta)$, $G_r(z_j|z_{j-1}; \theta)$, $\ell_s(\theta; z_{j-1}, z_j)$, $j = 1, \ldots, i$, and $G_{rs}(z_i|z_{i-1}; \theta)$ are defined as in Section 2.

The correction terms $h_i(z_i, y_n, z_{i-1}; \theta)$, $i = 1, \ldots, m$, are those ones needed for improving the univariate prediction limits $z^i_\alpha(\widehat{\theta}, z_{i-1})$; in particular, $h_1(z_1, y_n, z_0; \theta)$ corresponds to the term $Q(z, u, \theta)$, specified in Section 2, with $z = z_1$ and $u = y_n$. The additional terms $h_{ij}(z_i, z^{j+1}_{j-1}; \theta)$, $i = 2, \ldots, m$, $j = 1, \ldots, i-1$, correct for the part of the coverage error due to the additional

dependency among the quantiles $z_\alpha^i(\theta, z_{i-1})$, $i = 1, \ldots, m$, after estimating θ with the same estimator $\widehat{\theta}$.

Note that, both the estimative and the improved prediction regions do not necessarily follow a rectangular-type form. Honestly, this could represent a practical limitation for the present approach, since in the applications it is quite common to specify a multivariate forecast as a sequence of marginal prediction statements concerning each component of the future random vector Z. However, by considering rectangular prediction regions one implicitly assumes that the first realizations of the future path do not influence the positions and the width of the prediction intervals associated to the future subsequent observations. The simultaneous prediction regions studied in this paper overtake this limitation and they enable a suitable conditional predictive analysis, which may produce sequences of conditional prediction intervals related to alternative potential scenarios for the future observations. For example, one possibility is to consider, as a path forecast for the future realizations z, a vector of suitable point predictors for Z, obtained with respect to an optimality criterion.

Moreover, in this framework, it is not difficult to specify the multivariate conditional predictive density $\tilde{g}(z; y)$ which gives the system of improved simultaneous prediction limits (6) and (7) as quantiles. More precisely, $\tilde{g}(z; y)$ is such that

$$\int_{-\infty}^{\tilde{z}_\alpha^1} \int_{-\infty}^{\tilde{z}_\alpha^2} \cdots \int_{-\infty}^{\tilde{z}_\alpha^m} \tilde{g}(z; y)\, dz_m \cdots dz_1 = \alpha,$$

and, up to terms of order $O(n^{-1})$, it corresponds to

$$\begin{aligned}
\tilde{g}(z;y) &= g(z_1|y_n;\widehat{\theta})\left\{1+\frac{dh_1(z_1,y_n,z_0;\widehat{\theta})/dz_1}{g(z_1|y_n;\widehat{\theta})}\right\} \\
&\quad \prod_{i=2}^{m} g(z_i|z_{i-1};\widehat{\theta})\left\{1+\frac{dh_i(z_i,y_n,z_{i-1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}+\frac{\sum_{j=1}^{i-1} dh_{ij}(z_i,z_{j-1}^{j+1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}\right\} \\
&= g(z|y_n;\widehat{\theta})\left\{1+\sum_{i=1}^{m}\frac{dh_i(z_i,y_n,z_{i-1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}+\sum_{i=2}^{m}\frac{\sum_{j=1}^{i-1} dh_{ij}(z_i,z_{j-1}^{j+1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}\right\},
\end{aligned}$$

where $g(z|y_n;\widehat{\theta})$ and $g(z_i|z_{i-1};\widehat{\theta})$, $i=1,\ldots,m$, are, respectively, the joint conditional estimative predictive density of Z and the conditional estimative predictive density of Z_i given $Z_{i-1}=z_{i-1}$.

Furthermore, it is immediate to prove that, up to terms of order $O(n^{-1})$, the improved simultaneous prediction limits $\tilde{z}_\alpha^i$, $i=1,\ldots,m$, are such that

$$\int_{-\infty}^{\tilde{z}_\alpha^i} \tilde{g}(z_i;y,z_0^{i-1})\,dz_i=\alpha_i,$$

where the associated conditional predictive densities are

$$\tilde{g}(z_1;y)=g(z_1|y_n;\widehat{\theta})\left\{1+\frac{dh_1(z_1,y_n,z_0;\widehat{\theta})/dz_1}{g(z_1|y_n;\widehat{\theta})}\right\}$$

and, for $i=2,\ldots,m$,

$$\tilde{g}(z_i;y,z_1^{i-1})=g(z_i|z_{i-1};\widehat{\theta})\left\{1+\frac{dh_i(z_i,y_n,z_{i-1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}+\frac{\sum_{j=1}^{i-1} dh_{ij}(z_i,z_{j-1}^{j+1};\widehat{\theta})/dz_i}{g(z_i|z_{i-1};\widehat{\theta})}\right\}.$$

Finally, we emphasizes that, with an additional effort, it is possible to generalize these results to the case of Markov processes of order $k>1$. The system of improved simultaneous prediction limits is similar to that one defined above and it has the same structure. In particular, the correction terms for the univariate prediction limits correspond to $h^i(z_i,y_n,z_{i-1};\theta)$, $i=1,\ldots,m$, with $u=(y_{n-k+1},\ldots,y_n)$ substituted for y_n and z_{i-k}^{i-1} substituted for z_{i-1}, where $z_{i-h}=y_{n+i-h}$, $h\geq i$. On the other hand, the additional correction terms $h_{ij}(z_i,z_{j-1}^{j+1};\theta)$,

$i = 2, \ldots, m$, $j = 1, \ldots, i-1$, are suitably generalized in order to account for higher-order dependencies. In particular, we obtain that

$$h_{ij}(z_i, z_{j-k}^{i-1}; \theta) = \frac{1}{2}[2]i^{rs}(\theta) G_r(z_i | z_{i-k}^{i-1}; \theta) \left[\frac{\frac{d}{dz_j}\{G_s(z_j | z_{j-k}^{j-1}; \theta) g(z_{j+1} | z_{j-k+1}^{j}; \theta) \cdots g(z_{j+k} | z_j^{j+k-1}; \theta)\}}{g(z_j | z_{j-k}^{j-1}; \theta) \cdots g(z_{j+k} | z_j^{j+k-1}; \theta)} \right],$$

for $j = 1, \ldots, i-k-1$, $i = k+2, \ldots, m$,

$$h_{ij}(z_i, z_{j-k}^{i-1}; \theta) = \frac{1}{2}[2]i^{rs}(\theta) \left[\frac{\frac{d}{dz_j}\{G_s(z_j | z_{j-k}^{j-1}; \theta) g(z_{j+1} | z_{j-k+1}^{j}; \theta) \cdots g(z_{i-1} | z_{i-2}^{i-k-1}; \theta) G_r(z_i | z_{i-k}^{i-1}; \theta)\}}{g(z_j | z_{j-k}^{j-1}; \theta) \cdots g(z_{i-1} | z_{i-2}^{i-k-1}; \theta)} \right],$$

for $j = i-k, \ldots, i-2$, $i = 2, \ldots, m$, and

$$h_{ij}(z_i, z_{j-k}^{i-1}; \theta) = \frac{1}{2}[2]i^{rs}(\theta) \left[\frac{\frac{d}{dz_j}\{G_s(z_j | z_{j-k}^{j-1}; \theta) G_r(z_i | z_{i-k}^{i-1}; \theta)\}}{g(z_j | z_{j-k}^{j-1}; \theta)} \right],$$

for $j = i-1$, $i = 2, \ldots, m$.

The results presented in this section are undoubtedly relevant from a theoretical perspective since they apply to Markov process models the general findings of Corcuera and Giummolè (2006) on multivariate prediction. Moreover, they throw light on the structure of the required corrections for the simultaneous estimative prediction limits. However, the practical use of the improved prediction regions, with particular concern to the case of Markov processes of order $k > 1$, require too complex calculations. In order to overcome these computational difficulties a simulation based approach, obtained by extending to simultaneous prediction the Ueki and Fueda's (2007) procedure, can be defined. Vidoni (2015) is a first attempt in this direction.

4 Simultaneous prediction limits for general autoregressive models

In this section, the results on improved simultaneous prediction limits are applied in the following general framework already introduced by Vidoni (2009) for univariate prediction. Let

us considered the class of autoregressive time series models $\{Y_n\}_{n\geq 1}$ defined as

$$Y_n = \mu_n + \sigma_n \varepsilon_n, \quad n \geq 1,$$

with $\{\varepsilon_n\}_{n\geq 1}$ a sequence of independent, identically distributed, random variables having $E(\varepsilon_n) = 0$ and $V(\varepsilon_n) = 1$, $n \geq 1$. We state that, for each $n \geq 1$,

$$\begin{aligned} \mu_n &= \mu_n(Y_{n-1}, \ldots, Y_{n-p}; \beta), \\ \sigma_n &= \sigma_n(Y_{n-1}, \ldots, Y_{n-q}; \gamma) \end{aligned}$$

depend on past $p \geq 0$ and $q \geq 0$ observations, respectively, with $\theta = (\beta, \gamma)$ the unknown d-dimensional parameter. Moreover, we assume conditions assuring the existence of a stationary solution and we perform likelihood inference conditioned on $(Y_0, \ldots, Y_{-k+1}) = (y_0, \ldots, y_{-k+1})$, with $(y_0, \ldots, y_{-k+1})$ assumed to be known; $\widehat{\beta}$ and $\widehat{\gamma}$ indicate the estimators for β and γ.

Note that $\{Y_n\}_{n\geq 1}$ is a Markov process of order $k = \max(p, q)$, which includes the p-th order AR model and q-th order ARCH model, as well as the AR models with ARCH errors. With regard to the common distribution of the random variables ε_n, $n \geq 1$, we may consider the standard Gaussian distribution or alternative distributions with an additional shape parameter, such as the double gamma distribution and suitable heavy-tailed distributions.

In this framework, we have that, for $i = 1 \ldots, m$,

$$g(z_i | z_{i-k}^{i-1}; \theta) = \frac{1}{\sigma_{n+i}} f\left(\frac{z_i - \mu_{n+i}}{\sigma_{n+i}}\right), \quad G(z_i | z_{i-k}^{i-1}; \theta) = F\left(\frac{z_i - \mu_{n+i}}{\sigma_{n+i}}\right),$$

where μ_{n+i} and σ_{n+i}^2 are the conditional expectation and variance of Z_i given $Z_{i-k}^{i-1} = z_{i-k}^{i-1}$ and $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions of ε_n, $n \geq 1$.

In order to compute the improved simultaneous prediction limits (6) and (7), the partial derivatives of $G(z_i | z_{i-k}^{i-1}; \theta)$, $i = 1, \ldots, m$, with respect to components of β and γ, are required.

In particular, the first partial derivatives correspond to

$$\begin{aligned} G_{\beta,r}(z_i|z_{i-k}^{i-1};\theta) &= \frac{\partial G(z_i|z_{i-k}^{i-1};\theta)}{\partial \beta_r} = -f(u_i)\,\frac{\mu_{n+i,r}}{\sigma_{n+i}}, \\ G_{\gamma,r}(z_i|z_{i-k}^{i-1};\theta) &= \frac{\partial G(z_i|z_{i-k}^{i-1};\theta)}{\partial \gamma_r} = -u_i f(u_i)\,\frac{\sigma^2_{n+i,r}}{2\sigma^2_{n+i}}, \end{aligned}$$

with $u_i = (z_i - \mu_{n+i})/\sigma_{n+i}$, $\mu_{n+i,r} = \partial\mu_{n+i}/\partial\beta_r$ and $\sigma^2_{n+i,r} = \partial\sigma^2_{n+i}/\partial\gamma_r$. Indeed, the second partial derivatives are

$$\begin{aligned} G_{\beta\beta,rs}(z_i|z_{i-k}^{i-1};\theta) &= \frac{\partial^2 G(z_i|z_{i-k}^{i-1};\theta)}{\partial \beta_r \partial \beta_s} = f'(u_i)\,\frac{\mu_{n+i,r}\,\mu_{n+i,s}}{\sigma^2_{n+i}} - f(u_i)\,\frac{\mu_{n+i,rs}}{\sigma_{n+i}}, \\ G_{\beta\gamma,rs}(z_i|z_{i-k}^{i-1};\theta) &= \frac{\partial^2 G(z_i|z_{i-k}^{i-1};\theta)}{\partial \beta_r \partial \gamma_s} = \{f(u_i) + u_i f'(u_i)\}\,\frac{\mu_{n+i,r}\,\sigma^2_{n+i,s}}{2\sigma^3_{n+i}}, \\ G_{\gamma\gamma,rs}(z_i|z_{i-k}^{i-1};\theta) &= \frac{\partial^2 G(z_i|z_{i-k}^{i-1};\theta)}{\partial \gamma_r \partial \gamma_s} = \{3u_i f(u_i) + u_i^2 f'(u_i)\}\frac{\sigma^2_{n+i,r}\sigma^2_{n+i,s}}{4\sigma^4_{n+i}} - u_i f(u_i)\frac{\sigma^2_{n+i,rs}}{2\sigma^2_{n+i}}, \end{aligned}$$

where $f'(u_i) = df(u_i)/du_i$ and $\mu_{n+i,rs} = \partial^2\mu_{n+i}/\partial\beta_r\beta_s$, $\sigma^2_{n+i,rs} = \partial^2\sigma^2_{n+i}/\partial\gamma_r\gamma_s$.

Moreover, the first partial derivatives of $\log g(z_i|z_{i-k}^{i-1};\theta)$, $i = 1,\ldots,m$, with respect to components of β and γ, correspond to

$$\begin{aligned} \ell_{\beta,r}(\theta; z_{i-k}^{i-1}, z_i) &= \frac{\partial \log g(z_i|z_{i-k}^{i-1};\theta))}{\partial \beta_r} = -\frac{f'(u_i)}{f(u_i)}\,\frac{\mu_{n+i,r}}{\sigma_{n+i}}, \\ \ell_{\gamma,r}(\theta; z_{i-k}^{i-1}, z_i) &= \frac{\partial \log g(z_i|z_{i-k}^{i-1};\theta))}{\partial \gamma_r} = -\left\{u_i\frac{f'(u_i)}{f(u_i)} + 1\right\}\,\frac{\sigma^2_{n+i,r}}{2\sigma^2_{n+i}} \end{aligned}$$

and, finally, $dG_r(z_i|z_{i-k}^{i-1};\theta)/dz_h$ and $dg(z_i|z_{i-k}^{i-1};\theta)/dz_h$, with $h = i-k,\ldots,i-1$, may be obtained by derivation, with respect to z_h, of the functions $G_{\beta,r}(z_i|z_{i-k}^{i-1};\theta)$, $G_{\gamma,r}(z_i|z_{i-k}^{i-1};\theta)$ and $g(z_i|z_{i-k}^{i-1};\theta)$ specified before.

Whenever we assume a standard Gaussian distribution for the random variables ε_n, $n \geq 1$, the above calculations are simplified, since, in this particular case, $f'(u_i) = -u_i f(u_i)$.

4.1 An application to AR(1) models

Let us consider a stationary first-order Gaussian AR process $\{Y_n\}_{n\geq 1}$ defined as

$$Y_n = \mu + \rho\,(Y_{n-1} - \mu) + \varepsilon_n, \quad n \geq 1,$$

where $\mu \in \mathbb{R}$, $|\rho| < 1$ are unknown parameters and $\{\varepsilon_n\}_{n\geq 1}$ is a sequence of independent normal distributed random variables with zero mean and unknown variance σ^2. With the usual notation, $\theta = (\theta_1, \theta_2, \theta_3) = (\beta_1, \beta_2, \gamma_1) = (\mu, \rho, \sigma^2)$ and the transitive statistic is $U = Y_n$. Indeed, likelihood inference is conditioned on $Y_0 = y_0$, with y_0 known, and it is easy to see that Z_i given $Z_{i-1} = z_{i-1}$, $i = 1, \ldots, m$, follows a normal distribution with mean $\mu_{n+i} = \mu + \rho\,(z_{i-1} - \mu)$ and variance $\sigma^2_{n+i} = \sigma^2$. Thus, $g(z_i|z_{i-1};\theta) = \sigma^{-1}\phi\{(z_i - \mu_{n+i})/\sigma\}$ and $G(z_i|z_{i-1};\theta) = \Phi\{(z_i - \mu_{n+i})/\sigma\}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the density and the distribution function of a standard normal random variable.

For this simple stationary time series model, the expected information matrix is such that $i_{11}(\theta) = n(1-\rho)^2/\sigma^2$, $i_{22}(\theta) = n/(1-\rho^2) + o(n)$, $i_{33}(\theta) = n/(2\sigma^4)$ and $i_{rs}(\theta) = 0$, for $r \neq s$, and the maximum likelihood estimators are explicitly known and they correspond to

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

with $\widehat{\mu}_i = \widehat{\mu} + \widehat{\rho}\,(Y_{i-1} - \widehat{\mu})$. Indeed, to the relevant order of approximation, the bias of $\widehat{\theta}$, conditional on $Y_n = y_n$, is

$$\begin{aligned}
b_1(\theta, y_n) &= \frac{(y_n - \mu)(1+\rho)}{n(1-\rho)}, \\
b_2(\theta, y_n) &= \frac{(y_n - \mu)^2(1-\rho^2)\rho}{n\sigma^2} - \frac{1+4\rho}{n}, \\
b_3(\theta, y_n) &= \frac{(y_n - \mu)^2(1-\rho^2)}{n} - \frac{3\sigma^2}{n}.
\end{aligned}$$

In this context, the system of simultaneous estimative prediction limits is readily available and it corresponds to $z_\alpha^i(\widehat{\theta}, z_{i-1}) = \widehat{\mu}_{n+i} + q_i\widehat{\sigma}$, $i = 1, \ldots, m$, with $\widehat{\mu}_{n+i} = \widehat{\mu} + \widehat{\rho}\,(z_{i-1} - \widehat{\mu})$ and q_i such that $\Phi(q_i) = \alpha_i$. For the specification of the improved simultaneous prediction limits, the following quantities, obtained from the general formulae given above, are required. Since $\phi'(u) = -u\phi(u)$, we have that $G_1(z_i|z_{i-1};\theta) = -\phi(u_i)(1-\rho)/\sigma$, $G_2(z_i|z_{i-1};\theta) = -\phi(u_i)(z_{i-1} - \mu)/\sigma$, $G_3(z_i|z_{i-1};\theta) = -\phi(u_i)\,u_i/(2\sigma)$ and

$$
\begin{aligned}
G_{11}(z_i|z_{i-1};\theta) &= -\phi(u_i)\,u_i(1-\rho)^2/\sigma^2,\\
G_{12}(z_i|z_{i-1};\theta) &= -\phi(u_i)\{u_i(1-\rho)(z_{i-1}-\mu) - \sigma\}/\sigma^2,\\
G_{13}(z_i|z_{i-1};\theta) &= -\phi(u_i)(u_i^2-1)(1-\rho)/(2\sigma^3),\\
G_{22}(z_i|z_{i-1};\theta) &= -\phi(u_i)\,u_i(z_{i-1}-\mu)^2/\sigma^2,\\
G_{23}(z_i|z_{i-1};\theta) &= -\phi(u_i)(u_i^2-1)(z_{i-1}-\mu)/(2\sigma^3),\\
G_{33}(z_i|z_{i-1};\theta) &= -\phi(u_i)(u_i^3-3u_i)/(4\sigma^4),
\end{aligned}
$$

with $u_i = (z_i - \mu_{n+i})/\sigma$. Moreover, for $j = 1, \ldots, i-2$, $d\log g(z_{j+1}|z_j;\theta)/dz_j = u_{j+1}\rho/\sigma$ and, for $j = 1, \ldots, i$, $\ell_1(\theta; z_{j-1}, z_j) = u_j(1-\rho)/\sigma$, $\ell_2(\theta; z_{j-1}, z_j) = u_j(z_{j-1}-\mu)/\sigma$, $\ell_3(\theta; z_{j-1}, z_j) = (u_j^2-1)/(2\sigma^2)$, with $u_j = (z_j - \mu_{n+j})/\sigma$. Finally, we obtain that

$$
\begin{aligned}
\frac{dG_1(z_i|z_{i-1};\theta)}{dz_{i-1}} &= -\phi(u_i)\,u_i\frac{\rho(1-\rho)}{\sigma},\\
\frac{dG_2(z_i|z_{i-1};\theta)}{dz_{i-1}} &= -\phi(u_i)\,\frac{u_i\,\rho(z_{i-1}-\mu)+\sigma}{\sigma^2},\\
\frac{dG_3(z_i|z_{i-1};\theta)}{dz_{i-1}} &= -\phi(u_i)\,(u_i^2-1)\frac{\rho}{2\sigma^3}.
\end{aligned}
$$

With some algebra, we find that the simultaneous improved prediction limits (6) and (7)

are given by

$$\tilde{z}_\alpha^1(y) = \widehat{\mu}_{n+1} + \widehat{\sigma} q_1 + \frac{3\widehat{\rho}(y_n - \widehat{\mu})}{n} - \frac{(y_n - \widehat{\mu})^3(1-\widehat{\rho}^2)\widehat{\rho}}{n\widehat{\sigma}^2} + \widehat{\sigma}\left(\frac{9q_1}{4n} + \frac{q_1^3}{4n}\right)$$

and, for $i = 2, \ldots, m$, by

$$\begin{aligned}
\tilde{z}_\alpha^i(y, z_1^{i-1}) &= \widehat{\mu}_{n+i} + \widehat{\sigma} q_i - \frac{(y_n - \widehat{\mu})(1+\widehat{\rho}) - (z_{i-1} - \widehat{\mu})(1+4\widehat{\rho}) + (1-\widehat{\rho}^2)(z_{i-2} - \widehat{\mu})}{n} \\
&- \frac{(y_n - \widehat{\mu})^2(z_{i-1} - \widehat{\mu})(1-\widehat{\rho}^2)\widehat{\rho}}{n\widehat{\sigma}^2} + \frac{\widehat{\sigma}}{n}\sum_{j=1}^{i-2}(\widehat{u}_j - \widehat{u}_{j+1}\widehat{\rho}) \\
&+ \frac{(z_{i-1} - \widehat{\mu})(1-\widehat{\rho}^2)}{\widehat{\sigma} n}\sum_{j=1}^{i-2}(z_{j-1} - \widehat{\mu})(\widehat{u}_j - \widehat{u}_{j+1}\widehat{\rho}) \\
&+ \frac{\widehat{\sigma} q_i}{2n}\left\{\frac{(z_{i-1} - \widehat{\mu})^2(1-\widehat{\rho}^2)}{\widehat{\sigma}^2} - \frac{(y_n - \widehat{\mu})^2(1-\widehat{\rho}^2)}{\widehat{\sigma}^2} + \frac{9}{2} - 2\widehat{\rho}\right. \\
&- \left.\frac{2\widehat{\rho}(1-\widehat{\rho}^2)(z_{i-1} - \widehat{\mu})(z_{i-2} - \widehat{\mu})}{\widehat{\sigma}^2} + \sum_{j=1}^{i-2}(\widehat{u}_j^{\,2} - 1 - \widehat{u}_j\widehat{u}_{j+1}\widehat{\rho})\right\} \\
&- (q_i^2 - 1)\frac{\widehat{\sigma}\widehat{\rho}\widehat{u}_{i-1}}{2n} + \frac{\widehat{\sigma} q_i^3}{4n},
\end{aligned}$$

with $\widehat{u}_j = (z_j - \widehat{\mu}_{n+j})/\widehat{\sigma}$, $j = 1, \ldots, i-1$. For the case $i = 2$, the summations in the above formula have to be set equal to zero.

Furthermore, the conditional predictive densities giving $\tilde{z}_\alpha^i(y)$, $i = 1, \ldots, m$, as α_i-quantiles are, to the relevant order of approximation,

$$\begin{aligned}
\tilde{g}(z_1; y) &= \frac{1}{\widehat{\sigma}}\,\phi(\widehat{u}_1)\left[1 - \frac{\widehat{u}_1}{n}\left\{\frac{(y_n - \widehat{\mu})^3(1-\widehat{\rho}^2)\widehat{\rho}}{\widehat{\sigma}^3} - \frac{3\widehat{\rho}(y_n - \widehat{\mu})}{\widehat{\sigma}}\right\}\right. \\
&+ \left.\frac{3(\widehat{u}_1^2 - 1)}{n} + \frac{\widehat{u}_1^4 - 6\widehat{u}_1^2 + 3}{4n}\right]
\end{aligned}$$

and, for $i = 2, \ldots, m$,

$$\begin{aligned}
\tilde{g}(z_i; y, z_1^{i-1}) &= \frac{1}{\widehat{\sigma}}\phi(\widehat{u}_i)\left[1 + \frac{\widehat{u}_i}{n}\left\{\frac{(y_n - \widehat{\mu})(1+\widehat{\rho}) - (z_{i-1} - \widehat{\mu})(\widehat{\rho}^2 + 4\widehat{\rho})}{\widehat{\sigma}}\right.\right. \\
&\quad \left. + \frac{(y_n - \widehat{\mu})^2(z_{i-1} - \widehat{\mu})(1-\widehat{\rho}^2)\widehat{\rho}}{\widehat{\sigma}^3}\right\} + \frac{\widehat{u}_i^4 - 6\widehat{u}_i^2 + 3 - \widehat{\rho}\widehat{u}_{i-1} + 3\widehat{\rho}\widehat{u}_i\widehat{u}_{i-1}}{4n} \\
&+ \frac{\widehat{u}_i}{n}\left\{\sum_{j=1}^{i-2}(\widehat{u}_j - \widehat{u}_{j+1}\widehat{\rho}) + \frac{(1-\widehat{\rho}^2)(z_{i-1} - \widehat{\mu})}{\widehat{\sigma}^2}\sum_{j=1}^{i-2}(z_{j-1} - \widehat{\mu})(\widehat{u}_j - \widehat{u}_{j+1}\widehat{\rho})\right. \\
&+ \frac{\widehat{u}_i^2 - 1}{2n}\left\{\sum_{j=1}^{i-2}(\widehat{u}_j^2 - 1 - \widehat{u}_j\widehat{u}_{j+1}\widehat{\rho}) - \frac{2\widehat{\rho}(1-\widehat{\rho}^2)(z_{i-1} - \widehat{\mu})(z_{i-2} - \widehat{\mu})}{\widehat{\sigma}^2}\right. \\
&\quad \left.\left. + \frac{(z_{i-1} - \widehat{\mu})^2(1-\widehat{\rho}^2) - (y_n - \widehat{\mu})^2(1-\widehat{\rho}^2)}{\widehat{\sigma}^2} - \frac{\widehat{\rho}\widehat{u}_{i-1}}{2} + 6\right\}\right].
\end{aligned}$$

As emphasized before, for the case $i = 2$, the summations have to be set equal to zero. Notice that $\tilde{z}_\alpha^1(y)$ corresponds to the improved univariate prediction limit for $Z = Y_{n+1}$ and $\tilde{g}(z_1; y)$ is the associated predictive density.

We conclude this section by presenting the results of a simple simulation study where we compare the performance of the improved solution with that of the estimative one, considering a particular AR(1) model. We shall estimate the conditional coverage probabilities for the estimative and the improved simultaneous prediction limits of level $\alpha = 0.9, 0.95$ using the algorithm proposed by Kabaila (1999b). The estimates of the conditional coverage probabilities are based on 10,000 samples of dimension $n = 50, 100, 200$ 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) $\rho = 0.9$. The simultaneous prediction regions have dimension $m = 5, 10, 25$ and $\alpha_i = \alpha^{1/m}$, $i = 1, \ldots, m$. Similar results are obtained with alternative values for the observations y_n, y_0 and for the model parameters.

The results, presented in Table 1, show that the modified simultaneous prediction limits remarkably improve on the estimative ones. The improvement is more pronounced when $\rho = 0.9$

and it is more evident when the dimension m of the future random vector increases. Thus, the theoretical findings presented in the preceding sections turn out to be confirmed by this simple Monte Carlo experiment, which emphasized that the modified simultaneous prediction limits constitute a valuable improvement over the simple estimative solution.

Table 1 here

Finally, Figure 1 shows the plots of bivariate prediction regions, having nominal coverage $\alpha = 0.9$, based respectively on the estimative and the improved simultaneous prediction limits for the future observations Z_1 and Z_2. These regions are computed using simulated samples of dimension $n = 50$ from an AR(1) Gaussian model with $y_0 = 0$, $\mu = 1$, $\sigma^2 = 1$ and $\rho = 0.5, 0.9$ and they are built by means of α_i-simultaneous symmetric prediction intervals with $\alpha_i = \alpha^{1/2}$, $i = 1, 2$. As expected, their aspect deviates from a rectangular-type form and the improved prediction regions are wider than the estimative ones, in order to consider the additional variability induced by the plug-in procedure and to reduce, consequently, the coverage error.

Figure 1 here

4.2 An application to ARCH(1) models

Let us consider a stationary first-order Gaussian ARCH process $\{Y_n\}_{n\geq 1}$ defined as

$$Y_n = \sqrt{\gamma_1 + \gamma_2 \, Y_{n-1}^2} \, \varepsilon_n, \quad n \geq 1,$$

where γ_1 and γ_2 are unknown parameters and $\{\varepsilon_n\}_{n\geq 1}$ is a sequence of independent standard Gaussian random variables. We assume $\gamma_1 > 0$ and $\gamma_2 \in [0, 3.56]$ to ensure strict stationarity.

The unknown parameter is $\theta = (\theta_1, \theta_2) = (\gamma_1, \gamma_2)$ and likelihood inference is conditioned on $Y_0 = y_0$, with y_0 known. Indeed, $U = Y_n$ is the transitive statistic and Z_i given $Z_{i-1} = z_{i-1}$, $i = 1, \ldots, m$, follows a normal distribution with mean $\mu_{n+i} = 0$ and variance $\sigma^2_{n+i} = \gamma_1 + \gamma_2 z^2_{i-1}$. Thus, $g(z_i|z_{i-1};\theta) = \sigma^{-1}_{n+i}\phi(z_i/\sigma_{n+i})$ and $G(z_i|z_{i-1};\theta) = \Phi(z_i/\sigma_{n+i})$.

For ARCH models, maximum likelihood estimators $\widehat{\theta}$ are not explicitly known, as well as the associated expected information matrix. However, the parameter estimates can be computed easily using numerical procedures and, if the maximization routine uses second-order derivatives, minus the inverse of the empirical Hessian provides a convenient estimation for the expected information matrix. With regard to the conditional bias terms $b_1(\theta, y_n)$ and $b_2(\theta, y_n)$, we can consider simulation-based estimates based on the method given by Kabaila (1999b). Whenever these quantities are provided, we can readily specify the system of simultaneous estimative prediction limits $z^i_\alpha(\widehat{\theta}, z_{i-1}) = q_i\widehat{\sigma}_{n+i}$, $i = 1, \ldots, m$, with $\widehat{\sigma}_{n+i} = \widehat{\gamma}_1 + \widehat{\gamma}_2 z^2_{i-1}$ and q_i such that $\Phi(q_i) = \alpha_i$.

Furthermore, the simultaneous improved prediction limits $\tilde{z}^1_\alpha(y)$ and $\tilde{z}^i_\alpha(y, z^{i-1}_1)$, $i = 2, \ldots, m$, can be computed using (6) and (7), by considering the following quantities, obtained by specifying for an ARCH(1) model the formulae given at the beginning of the present section. That is, since $\phi'(u) = -u\phi(u)$, we have that $G_1(z_i|z_{i-1};\theta) = -(1/2)\sigma^{-1}_{n+i}\phi(z_i/\sigma_{n+i})(z_i/\sigma^2_{n+i})$, $G_2(z_i|z_{i-1};\theta) = G_1(z_i|z_{i-1};\theta)z^2_{i-1}$ and

$$\begin{aligned}
G_{11}(z_i|z_{i-1};\theta) &= -\frac{1}{4\sigma_{n+i}}\phi\left(\frac{z_i}{\sigma_{n+i}}\right)\left(\frac{z_i^3}{\sigma^6_{n+i}} - 3\frac{z_i}{\sigma^4_{n+i}}\right), \\
G_{12}(z_i|z_{i-1};\theta) &= G_{11}(z_i|z_{i-1};\theta)z^2_{i-1}, \\
G_{22}(z_i|z_{i-1};\theta) &= G_{11}(z_i|z_{i-1};\theta)z^4_{i-1}.
\end{aligned}$$

Moreover, for $j = 1, \ldots, i-2$, $d\log g(z_{j+1}|z_j;\theta)/dz_j = \{(z_{j+1}/\sigma^4_{n+j+1}) - (1/\sigma^2_{n+j+1})\}\gamma_2 z_j$ and,

for $j = 1, \ldots, i$, $\ell_1(\theta; z_{j-1}, z_j) = (1/2)\{(z_j^2/\sigma_{n+j}^4) - (1/\sigma_{n+j}^2)\}$, $\ell_2(\theta; z_{j-1}, z_j) = \ell_1(\theta; z_{j-1}, z_j) z_{j-1}^2$. Finally, we get

$$\begin{aligned}
\frac{dG_1(z_i|z_{i-1};\theta)}{dz_{i-1}} &= \frac{1}{2\sigma_{n+i}}\phi\left(\frac{z_i}{\sigma_{n+i}}\right)\left(3\frac{z_i}{\sigma_{n+i}^4} - \frac{z_i^3}{\sigma_{n+i}^6}\right)\gamma_2 z_{i-1}, \\
\frac{dG_2(z_i|z_{i-1};\theta)}{dz_{i-1}} &= \frac{dG_1(z_i|z_{i-1};\theta)}{dz_{i-1}} z_{i-1}^2 + G_1(z_i|z_{i-1};\theta) z_{i-1}.
\end{aligned}$$

In addition to the simultaneous improved prediction limits, using the general formulae specified in Section 3, we can also find the conditional predictive densities giving $\tilde{z}_\alpha^i(y)$, $i = 1, \ldots, m$, as α_i-quantiles to the relevant order of approximation.

Finally, a simple simulation study compares the conditional coverage probability of the estimative and the modified simultaneous prediction limits. Samples of size $n = 50, 100, 200$ are generated from an ARCH(1) model with $\gamma_1 = 0.5$ and (a) $\gamma_2 = 0.5$, (b) $\gamma_2 = 1$ and with the last and the first observations fixed to $y_n = 1$ and $y_0 = 0$, respectively. The simultaneous prediction regions have dimension $m = 5, 10, 25$ and $\alpha_i = \alpha^{1/m}$, $i = 1, \ldots, m$. Similar results are obtained with alternative values for the observations y_n, y_0 and for the model parameters. The simulations are based on 5,000 replications and the asymptotic bias of the maximum likelihood estimators is estimated using the Kabaila's simulation-based procedure with 1,000 additional parametric bootstrap samples. The results are shown in Table 2 and emphasize that the improved procedure performs uniformly better than that one based on the estimative solution, even if there is a mild tendency, for the improve method, to overestimate the coverage probability when the dimension m of the future random vector increases.

Table 2 here

As for the AR models, it is easy to plot the bivariate prediction regions based, respectively,

on the estimative and the improved simultaneous prediction limits for the future observations Z_1 and Z_2. These regions are are built by considering α_i-simultaneous symmetric prediction intervals with $\alpha_i = \alpha^{1/2}$, $i = 1, 2$, and $\alpha = 0.9$. Indeed, they are computed using simulated samples of dimension $n = 50$ from an ARCH(1) model with $y_0 = 0$, $\gamma_1 = 0.5$ and $\gamma_2 = 0.5, 1$. As shown in Figure 2, the improved prediction regions are wider than the estimative ones, in order to account for the additional variability induced by the estimative procedure, and their aspect definitely deviates from a rectangular-type form, displaying the fact that the conditional variance is defined as a quadratic function of the past observation.

Figure 2 here

References

[1] Barndorff-Nielsen OE, Cox DR (1996) Prediction and asymptotics. Bernoulli 2:319–340.

[2] Beran R (1990) Calibrating prediction regions. J Amer Statist Assoc 85:715–723.

[3] Bjørnstad JF (1996) On the generalization of the likelihood function and the likelihood principle. J Amer Statist Assoc 91:791–806.

[4] Concordet D, Servien R (2014) Individual prediction regions for multivariate longitudinal data with small samples. Biometrics 70:629–638.

[5] Corcuera JM (2008) Approximate predictive pivots for autoregressive processes. Statist Probab Lett 78:2685–2691.

[6] Corcuera JM, Giummolè F (2006) Multivariate prediction. Bernoulli 12:157–168.

[7] Hall P, Peng L, Tajvidi N (1999) On prediction intervals based on predictive likelihood or bootstrap methods. Biometrika 86:871–880.

[8] Jordà Ò, Marcellino M (2010) Path forecast evaluation. J Appl Econ 25:635–662.

[9] Kabaila P (1999a) The relevance property for prediction intervals. J Time Series Anal 20:655–662.

[10] Kabaila P (1999b) An efficient simulation method for the computation of a class of conditional expectations. Aust N Z J Stat 41:331–336.

[11] Kabaila P, He Z (2004) The adjustment of prediction intervals to account for errors in parameter estimation. J Time Series Anal 25:351–358.

[12] Kabaila P, Syuhada K (2007) Improved prediction limits for AR(p) and ARCH(p) processes. J Time Series Anal 29:213–223.

[13] Lawless JF, Fredette M (2005) Frequentist prediction intervals and predictive distributions. Biometrika 92:529–542.

[14] Scheffé H (1959) The Analysis of Variance. Wiley, New York.

[15] Ueki M, Fueda K (2007) Adjusting estimative prediction limits. Biometrika 94:509–511.

[16] Vidoni P (1998) A note on modified estimative prediction limits and distributions. Biometrika 85:949–953.

[17] Vidoni P (2004) Improved prediction intervals for stochastic process models. J Time Series Anal 25:137–154.

[18] Vidoni P (2009) A simple procedure for computing improved prediction intervals for autoregressive models. J Time Series Anal 30:577–590.

[19] Vidoni P (2015) Calibrated multivariate distributions for improved conditional prediction. J Multiv Anal 142:16–25.

[20] Wang WL, Fan TH (2010) ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. Comput Statist Data Anal 54:1328–1341.

[21] Wolf M, Wunderli D (2015) Bootstrap joint prediction regions. J Time Series Anal 36:352–376.

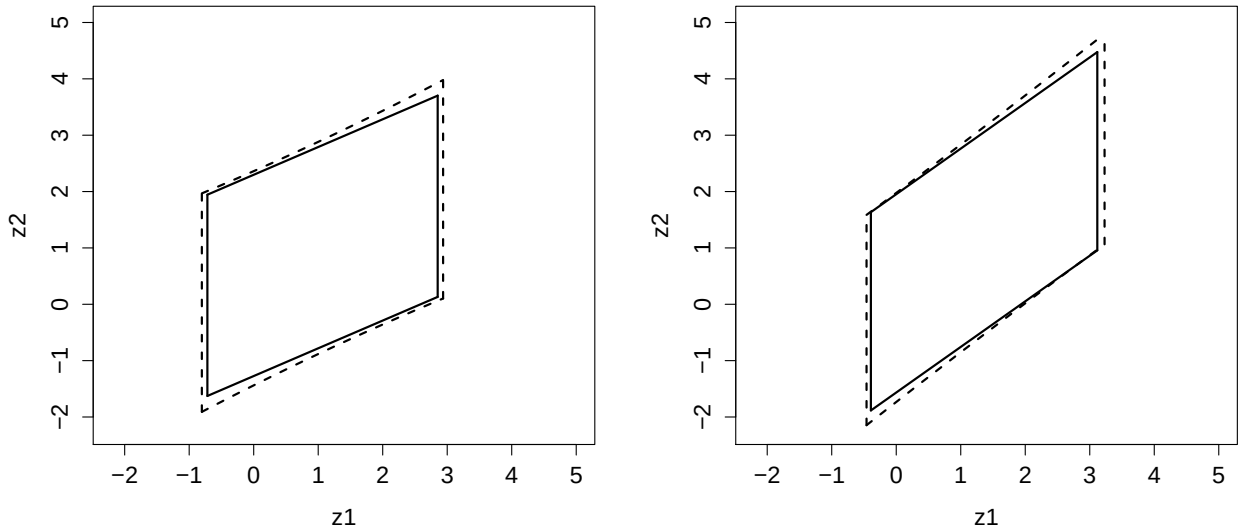

Figure 1: Estimative (solid line) and improved (dashed line) bivariate prediction regions with $\alpha = 0.9$, based on samples of dimension $n = 50$ from an AR(1) Gaussian model with $y_0 = 0$, $\mu = 1$, $\sigma^2 = 1$ and $\rho = 0.5$ (left), $\rho = 0.9$ (right).

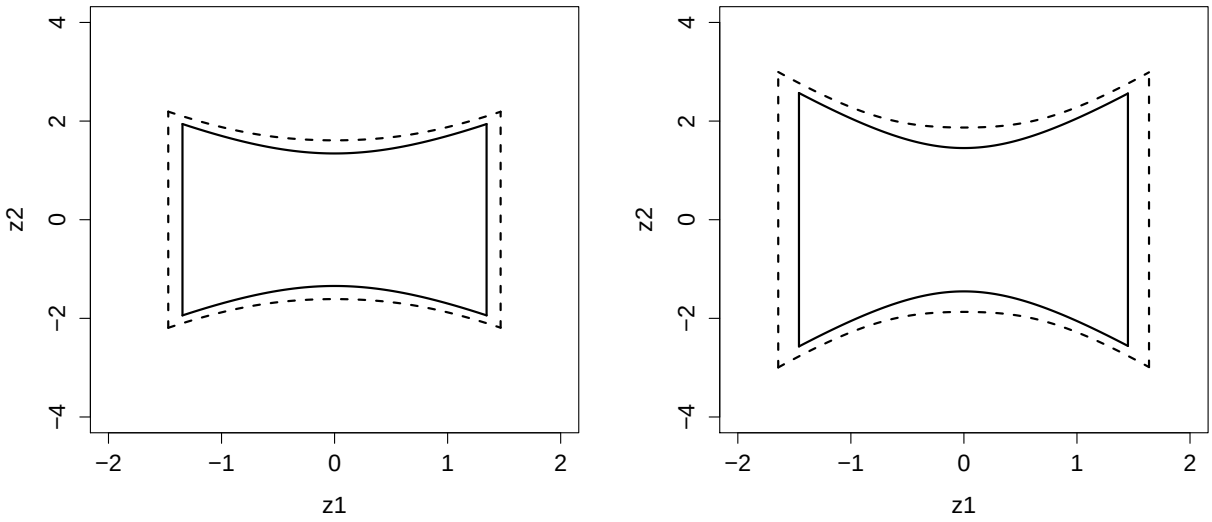

Figure 2: Estimative (solid line) and improved (dashed line) bivariate prediction regions with $\alpha = 0.9$, based on samples of dimension $n = 50$ from an ARCH(1) model with $y_0 = 0$, $\gamma_1 = 0.5$ and $\gamma_2 = 0.5$ (left), $\gamma_2 = 1$ (right).

			(a)		(b)	
α	n	m	Estimative	Improved	Estimative	Improved
0.9	50	5	0.858	0.889	0.841	0.868
		10	0.840	0.887	0.827	0.867
		25	0.813	0.891	0.783	0.843
	100	5	0.876	0.893	0.872	0.884
		10	0.870	0.892	0.867	0.883
		25	0.856	0.888	0.848	0.874
	200	5	0.890	0.896	0.890	0.895
		10	0.886	0.897	0.890	0.897
		25	0.877	0.893	0.877	0.888
0.95	50	5	0.916	0.939	0.930	0.941
		10	0.908	0.942	0.912	0.932
		25	0.887	0.945	0.854	0.910
	100	5	0.933	0.945	0.927	0.942
		10	0.928	0.943	0.914	0.933
		25	0.921	0.944	0.912	0.932
	200	5	0.944	0.951	0.934	0.947
		10	0.944	0.951	0.946	0.951
		25	0.933	0.942	0.933	0.942

Table 1: AR(1) Gaussian model with $\mu = 1$, $\sigma^2 = 1$ and (a) $\rho = 0.5$, (b) $\rho = 0.9$. Conditional coverage probabilities for estimative and improved simultaneous prediction limits of level $\alpha = 0.9, 0.95$, with $m = 5, 10, 25$. Estimation is based on 10,000 Monte Carlo conditional (on $y_n = 1$) samples of dimension $n = 50, 100, 200$, with $y_0 = 0$. Estimated standard errors are smaller than 0.0040, when $\alpha = 0.9$, and smaller than 0.0036, when $\alpha = 0.95$.

			(a)		(b)	
α	n	m	Estimative	Improved	Estimative	Improved
0.9	50	5	0.876	0.909	0.873	0.906
		10	0.866	0.926	0.847	0.904
		25	0.837	0.928	0.825	0.909
	100	5	0.893	0.912	0.884	0.904
		10	0.880	0.911	0.878	0.908
		25	0.867	0.917	0.861	0.907
	200	5	0.894	0.902	0.891	0.903
		10	0.892	0.906	0.889	0.904
		25	0.891	0.918	0.888	0.914
0.95	50	5	0.923	0.954	0.926	0.954
		10	0.913	0.961	0.915	0.954
		25	0.894	0.961	0.886	0.948
	100	5	0.936	0.953	0.935	0.953
		10	0.933	0.957	0.932	0.954
		25	0.930	0.966	0.927	0.956
	200	5	0.940	0.948	0.941	0.947
		10	0.945	0.955	0.943	0.952
		25	0.943	0.957	0.936	0.951

Table 2: ARCH(1) model with $\gamma_1 = 0.5$ and (a) $\gamma_2 = 0.5$, (b) $\gamma_2 = 1$. Conditional coverage probabilities for estimative and improved simultaneous prediction limits of level $\alpha = 0.9, 0.95$, with $m = 5, 10, 25$. Estimation is based on 5,000 Monte Carlo conditional (on $y_n = 1$) samples of dimension $n = 50, 100, 200$, with $y_0 = 0$. Estimated standard errors are smaller than 0.0054, when $\alpha = 0.9$, and smaller than 0.0045, when $\alpha = 0.95$.