Approximate predictive pivots for autoregressive processes

José M. Corcuera

Universitat de Barcelona, Facultat de Matemàtiques, Gran Via de les Corts, Catalanes 585, 08007 Barcelona, Spain

ABSTRACT

In this paper the author considers an autoregressive process where the parameters of the process are unknown and try to obtain pivots for predicting future observations. If we do a probabilistic prediction with the estimated model, where the parameters are estimated by a sample of size \( n \), we introduce an error of order \( n^{-1} \) in the coverage probabilities of the prediction intervals. However we can reduce the order of the error if we calibrate adequately the estimated prediction bounds. The solution obtained can be expressed in terms of an approximate predictive pivot.

1. Introduction

The general setting is of prediction of an absolutely continuous (a.c.) random variable \( Z \) based on the observation \( y = (y_1, y_2, \ldots, y_n) \) corresponding to a random vector \( Y = (Y_1, Y_2, \ldots, Y_n) \), where the laws of \( Y \) and \( Z \) depend on a common and unknown parameter \( \theta \in \Theta \subset \mathbb{R}^d \). A prediction statement about \( Z \) is often given by prediction limits, i.e. real functions \( K_\alpha(\cdot) \) such that

\[
P_\theta\{Z \leq K_\alpha(Y)\} = \alpha,
\]

for every \( \theta \in \Theta \) and for any fixed \( \alpha \in (0, 1) \). The above probability is usually called coverage probability and it is calculated with respect to the joint density of \( Z \) and \( Y \). Sometimes the existence of exact (predictive) pivotal quantities, that is of functions of \( Z \) and \( Y \) whose distribution does not depend on \( \theta \), permit us to find an exact solution. But this is the exception. Here we look for approximate prediction limits and predictive pivots. An approximate solution is to take \( K_\alpha(Y) = q_\alpha(\hat{\theta}) \), where \( q_\alpha(\cdot) \) is the \( \alpha \)-quantile of the conditional distribution of \( Z \) given \( Y = y \), that we also assume absolutely continuous, and \( \hat{\theta} \) is an efficient estimator of \( \theta \). Note that, if we denote the conditional density of \( Z \) given \( Y = y \), \( g(z; \theta|y) \), then \( q_\alpha(\hat{\theta}) \) will be the \( \alpha \)-quantile of the so-called estimative density \( g(z; \theta|y) \). However these predictions limits are usually imprecise, having coverage error of order \( O(n^{-1}) \), that is

\[
P_\theta\{Z \leq q_\alpha(\hat{\theta})\} = \alpha + O(n^{-1}).
\]

This is a well known result; indeed Barndorff-Nielsen and Cox (1996) suggest a way to correct these quantiles obtaining prediction limits with a coverage error of order \( o(n^{-1}) \). The solution can be expressed in terms of a predictive density whose quantiles are precisely these predictions bounds. We will apply this method to the case where \( Y = (Y_1, Y_2, \ldots, Y_n) \) is such that

\[
Y_{k+1} - \mu = \sum_{j=1}^{p} \delta_j (Y_{k+j+1} - \mu) + \varepsilon_{k+1}, \quad k \in \mathbb{Z},
\]
where the innovation \( \epsilon_{t+1} \sim N(0, \sigma^2) \), \( \phi_1, \ldots, \phi_p \) are the coefficients of a stationary autoregressive process of order \( p, AR(p) \) for short, and where \( Y_0, Y_1, \ldots, Y_{p+1} \) are assumed to be known and fixed. All the parameters, \( \mu, \phi_1, \ldots, \phi_p \) and \( \sigma \) are unknown and \( Z = Y_{p+1} \).

The paper is organized in the following way. In the next section we give the details of the Barndorff-Nielsen and Cox method. Then we solve the \( AR(p) \) case. Some simulations are made in order to study how the corrections work.

2. The method

In this section we explain, in a concise and heuristic manner, the method of Barndorff-Nielsen and Cox (1996) to modify the estimative density \( g(z; \hat{\theta}|y) \) to get better, asymptotically, prediction intervals or limits and how to obtain approximate predictive pivots. For more details, we refer the reader to that paper.

When we use \( g(z; \hat{\theta}|y) \) instead of \( g(z; \theta|y) \), we are introducing the error in the estimation of \( \theta \); we can correct this effect if we take into account the incertitude in \( \hat{\theta} \) about \( \theta \). We know, by the conditioning principle that, when estimating \( \theta \), we should consider the conditional distribution of \( \hat{\theta} \) given \( A(Y) = a \), where \( A \) is an ancillary statistics (distribution free of \( \theta \) and such that \( (\hat{\theta}, A) \) is a minimal sufficient statistics for \( \theta \)). Then, if we have a minimal predictive sufficient reduction of the data, \( W \), we take \( W \) instead of \( Y \), and if there is a decomposition \( W = (\hat{\theta}, A) \) where \( A \) has a distribution free of \( \theta \), or approximately free (here we use the continuity principle), we shall suppose \( A \) is fixed when we study long-run properties.

We shall omit \( A \) in the notation, whenever this does not produce confusion. Consider now that the joint law of \( (Z, Y) \) is absolutely continuous, with respect to the Lebesgue measure. Then, to predict \( Z \) we look for certain functions \( K^{(\alpha)}(\hat{\theta}) \) such that,

\[
P_\theta(\alpha \leq \hat{Z}^o|A = a) = \alpha \quad \forall \theta \in \Theta, \alpha \in (0, 1),
\]

up to order \( o_p(n^{-1}) \).

Let \( G(z; \theta|\hat{\theta}) = P_\theta(Z \leq z|W = (\hat{\theta}, a), \theta \)\], that is the distribution function of \( Z \) given \( Y = y \) (note that \( W \) is transitive) considered as function of \( z, \theta, \hat{\theta} \) and \( a \); and \( g(z; \theta|\hat{\theta}) \) the corresponding density function. Note that \( P_\theta(Z \leq z|A = a) = E_\theta(G(z; \theta|\hat{\theta})|A = a) \).

Let \( q^{(\alpha)}(\hat{\theta}) \) be such that

\[
G(q^{(\alpha)}(\hat{\theta}); \theta|\hat{\theta}) = \alpha, \quad \forall \theta \in \Theta.
\]

Suppose we can write

\[
E_\theta(G(q^{(\alpha)}(\hat{\theta}); \theta|\hat{\theta})) = \alpha + Q(q^{(\alpha)}(\theta), \theta) + o_p(n^{-1}),
\]

where \( Q(q^{(\alpha)}(\theta), \theta) = \frac{b(q^{(\alpha)}(\theta), \theta)}{n} \), and \( b(\cdot, \cdot) \) is a smooth function, then

\[
\int_{-\infty}^{q^{(\alpha)}(\hat{\theta})} \frac{g(z; \theta|\hat{\theta})}{C(q^{(\alpha)}(\theta), \theta)} dz = G(q^{(\alpha)}(\hat{\theta}); \theta|\hat{\theta}) - \int_{q^{(\alpha)}(\hat{\theta})}^{\infty} G(z; \theta|\hat{\theta}) dz
\]

\[= G(q^{(\alpha)}(\hat{\theta}); \theta|\hat{\theta}) - Q(q^{(\alpha)}(\theta), \theta) + o_p(n^{-1}),\]

where we assume that, conditioning on \( A = a \), \( \hat{\theta} - \theta = O_p(n^{-1/2}) \). Then by taking the expectation with respect to \( \hat{\theta} \) we obtain that

\[
K^{(\alpha)}(\hat{\theta}) = q^{(\alpha)}(\hat{\theta}) - \frac{Q(q^{(\alpha)}(\theta), \theta)}{g(q^{(\alpha)}(\theta); \theta|\hat{\theta})} = q^{(\alpha)}(\hat{\theta}) - \frac{Q(q^{(\alpha)}(\hat{\theta}), \hat{\theta})}{g(q^{(\alpha)}(\theta); \theta|\hat{\theta})} + o_p(n^{-1}),
\]

is a solution of \( (1) \). Also we could get a predictive density, denoted by \( \hat{p}(z|y) \), with these quantiles up to order \( n^{-1} \), that is

\[
\int_{-\infty}^{\hat{K}^{(\alpha)}(\hat{\theta})} \hat{p}(z|y) dz = \alpha + o_p(n^{-1}).
\]

To obtain it, consider the identity

\[
\int_{-\infty}^{\hat{K}^{(\alpha)}(\hat{\theta})} g(u; \theta|\hat{\theta}) du = \alpha,
\]

by doing the change of variable \( u = z + \frac{q^{(\alpha)}(\hat{\theta})}{g(z; \theta|\hat{\theta})} \), we obtain that

\[
\int_{-\infty}^{z_0} (1 + \hat{r}) g(z + \hat{\theta}; \theta|\hat{\theta}) dz = \alpha,
\]

where

\[
\hat{r} = \frac{Q(z; \hat{\theta})}{g(z; \hat{\theta})}, \quad \text{and} \quad z_0 = K^{(\alpha)}(\hat{\theta}) + o_p(n^{-1}).
\]
So we can write
\[
\hat{p}(z|y) = (1 + \partial_z f)g(z + \hat{r}; \hat{\theta}(\hat{\theta})).
\]
In some cases it is possible to do a one to one transformation of \(z\), involving \(\hat{\theta}\), that is, \(v = f(z, \hat{\theta})\), such that the new predictive density does not depend on \(y\). Then \(f(z, \hat{\theta})\) will be an approximate predictive pivot.

In the rest of the section we will show how to calculate \(G(z, \theta)\). In general a symbol with a hat will denote the same quantity as without hat but where we substitute \(\hat{\theta}\) for \(\theta\). We shall use the repeated index convention, this mean if an index appears as an upper and a lower index in a term, then summation over the index is implied. For instance
\[
(\hat{\theta} - \theta)^\prime \partial_i G = \sum_{i=1}^d (\hat{\theta} - \theta)^\prime \partial_i G \text{ where } d \text{ is the dimension of } \theta.
\]

Considering now \(G(q^{(\alpha)}(\hat{\theta}); \theta(\hat{\theta}))\) as a, smooth, function of \(\hat{\theta}\) and by doing its Taylor's expansion around \(\theta\), we have:
\[
\begin{align*}
G(q^{(\alpha)}(\hat{\theta}); \theta(\hat{\theta})) &= \alpha + (\hat{\theta} - \theta)^\prime (gq_i + \partial \theta) + \frac{1}{2} (\hat{\theta} - \theta)^\prime (\hat{\theta} - \theta)^\prime (gq_r + \partial \theta) + gq_r + G_{rs} \\& + g_r q_s + G_{rst} + O_p(n^{-3/2}),
\end{align*}
\]
where
\[
\begin{align*}
\hat{g} &= \partial g/\partial z, q^{(\alpha)}_i &= \partial q^{(\alpha)}/\partial \theta, \partial q^{(\alpha)}_r &= \partial^2 q^{(\alpha)}/\partial \theta^2, \\
G_{rs} &= \partial G(z; \theta)/\partial \theta^r \partial \theta^s = \partial^2 G(z; \theta)/\partial \theta^r \partial \theta^s
\end{align*}
\]
evaluated in \(z = q^{(\alpha)}(\theta)\) and \(\hat{\theta} = \theta\). The asymptotic behavior, under mild regularity conditions, of \(\hat{\theta}\), and smoothness of the log-likelihood function are sufficient to guarantee the behavior of the rest. We can eliminate the derivatives of \(q^{(\alpha)}\) by using
\[
G(q^{(\alpha)}(\theta); \theta(\hat{\theta})) = \alpha, \quad \forall \theta \in \Theta.
\]
By taking the derivative with respect to \(\theta\), we obtain
\[
gq_r + G_{rs} + G_{rst} = 0,
\]
where
\[
G_{rs} = \partial G(z; \theta)/\partial \theta^r |_{\theta = \theta},
\]
and
\[
{}^{\hat{g}} q^{(\alpha)}_r + g_r q^{(\alpha)}_s + g_{rs} q^{(\alpha)} + g_{rs} q^{(\alpha)} + G_{rs} + G_{rst} + G_{rst} + G_{rst} = 0.
\]
Then, we can write
\[
G(q^{(\alpha)}(\hat{\theta}); \theta(\hat{\theta})) = \alpha - (\hat{\theta} - \theta)^\prime G_{rs} + \frac{1}{2} (\hat{\theta} - \theta)^\prime (\hat{\theta} - \theta)^\prime (2g_r + \partial \theta) + G_{rs} - G_{rst} + [2G_{rst}] + O_p(n^{-3/2}).
\]
Here, Barndorff-Nielsen and Cox (1996), indicates the sum of two terms, obtained by permutation of indices, \(r\) and \(s\), involved. Then we have,
\[
E_{\theta}(G(q^{(\alpha)}(\hat{\theta}); \theta(\hat{\theta}))) = \alpha - b^\prime G_{rs} + \frac{1}{2} b^\prime [2g_r + \partial \theta] + G_{rs} - G_{rst} + [2G_{rst}] + O_p(n^{-3/2}),
\]
where \(r^s\) and \(b^\prime\) are the variance and bias, respectively, of the asymptotic distribution of \(\hat{\theta}\) given \(A = a\). Also, in the regular cases we are considering the variance and bias of asymptotic distribution of the mle can be calculated from the conditional likelihood, see Barndorff-Nielsen and Cox (1994). In fact, let \(l(\theta, \hat{\theta}, a)\) be the log-likelihood function for \(\theta\) based on \(y\) as depending on \(y\) through \(w = (\hat{\theta}, a)\) only, then if we consider the Taylor expansion of the score function, we have
\[
l_r(\hat{\theta}; \hat{\theta}) = l_r(\hat{\theta} - \theta)^\prime + \frac{1}{2} l_{rs}(\hat{\theta} - \theta)^\prime (\hat{\theta} - \theta)^\prime + O_p(n^{-1/2}),
\]
where
\[
l_r = \partial l/\partial \theta^r, l_{rs} = \partial^2 l/\partial \theta^r \partial \theta^s, l_{rst} = \partial^3 l/\partial \theta^r \partial \theta^s \partial \theta^t
\]
and the derivatives after the semicolon, the so-called sample derivatives, are calculated in \(\hat{\theta} = \theta\). Here we are assuming, as in the regular cases, that the derivatives are \(O_p(n)\) and \(\hat{\theta} - \theta = O_p(n^{-1/2})\). Note that the first term of the Taylor expansion vanishes because \(\hat{\theta}\) is the mle, in fact, by calculating the total derivative, we obtain the so called Bartlett relations between the derivatives: \(l_{rs} + l_{sr} = 0, \ldots\). Then, by taking the (conditional) expectation (given \(A = a\)) we obtain the expression (6)
\[
E_{\theta}(\hat{\theta} - \theta)^\prime = -\frac{1}{2} b^\prime g^\prime + O_p(n^{-3/2}).
\]
because the score has zero expectation and because \((i_{r,3}) = (i_{n}) = (i^{*})^{-1}\). This latter fact can be obtained by taking into account that
\[ E_{0}(l_{i}(\theta; \hat{\theta})) = E_{0}(-l_{i}(\theta; \hat{\theta})) , \]
and calculating both sides of the equality by products or derivatives, respectively, of \((4)\).

So,
\[ (i^{*}) = -(l_{r,3})^{-1} \quad b^{*} = -\frac{1}{2} \frac{i^{*}}{i_{l_{r,3}}} , \]
and, we can, finally, write
\[ Q(z; \theta) = \frac{1}{2} i^{*}(H_{z} - i^{*}l_{y,3}H_{r}) , \]
where
\[ H_{z} = [2]g^{-1}_{r,3}(G_{r,3} + G_{r,3}) - G_{r,3} - [2]G_{r,3} \quad H_{r} = -G_{r,3} , \]
and we substitute \(z\) for \(q^{(a)}(\theta)\).

3. The AR(p) case with all parameters unknown

Suppose that we observe \(y = (y_{1}, y_{2}, \ldots, y_{n})\) such that
\[ y_{k+1} - \mu = \sum_{j=1}^{p} \phi_{j}(y_{k-j+1} - \mu) + \varepsilon_{k+1}, \quad k \in \mathbb{Z} , \]
where the innovation \(\varepsilon_{k+1} \sim N(0, \sigma^{2})\), \(\phi_{1}, \ldots, \phi_{p}\) are the autoregressive coefficients of a stationary AR(p) process and where \(y_{0}, y_{-1}, \ldots, y_{-p+1}\) are assumed to be known and fixed. All the parameters, \(\mu, \phi_{1}, \ldots, \phi_{p}\) and \(\sigma\) are unknown. The purpose is to obtain an approximate predictive pivot that gives exact (to order \(n^{-1}\)) prediction limits for future observations after 1 step. In this case \(z = y_{n+1}\) and \(r\) can be written as (see Appendix for the details)
\[ r = \frac{\sigma}{n} \left( u + v \Delta - \frac{\Delta^{3}}{4} \right) , \]
where
\[ \Delta = \frac{y_{n+1} - \mu - \sum_{j=1}^{p} \phi_{j}(y_{n-j+1} - \mu)}{\sigma} . \]

Then, since
\[ \partial_{y_{n+1}} \hat{\hat{r}} = \frac{1}{n} \left( \hat{\hat{v}} - 3 \hat{\Delta}^{2} \right) , \]
\[ g(y_{n+1}; \hat{\theta}) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^{2}} \left( y_{n+1} - \hat{\mu} - \sum_{j=1}^{p} \hat{\phi}_{j}(y_{n-j+1} - \hat{\mu}) \right)^{2} \right\} = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \hat{\Delta}^{2} \right\} , \]
the predictive density
\[ \hat{p}(y_{n+1}|y_{1}, \ldots, y_{n}) = (1 + \partial_{y_{n+1}} \hat{\hat{r}})g(y_{n+1} + \hat{\hat{r}}; \hat{\theta}) \]
\[ = \left( 1 + \frac{1}{n} \left( \hat{\hat{v}} - 3 \hat{\Delta}^{2} \right) \right) \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \hat{\Delta} + \frac{1}{n} (\hat{\mu} + \hat{\hat{v}} - \frac{\hat{\Delta}^{3}}{4}) \right)^{2} \right\} \]
\[ \propto \exp \left\{ -\frac{\hat{\Delta}^{2}}{n} - \left( \frac{1}{2} + \frac{4 \hat{\hat{v}} + 3}{4n} \right) \hat{\Delta}^{2} + \frac{\hat{\Delta}^{4}}{4n} \right\} + o_{p}(n^{-1}) \]
\[ \propto \frac{1}{\left( 1 + \frac{\hat{\Delta}^{2}}{4n} \right)^{k_{2}}} + o_{p}(n^{-1}) , \]
where
\[ k_{1} = n + \frac{4 \hat{\hat{v}} + 3}{2} \]
\[ k_{2} = \frac{1}{n} \left( n + \frac{4 \hat{\hat{v}} + 3}{2} \right)^{2} - 1 \]
\[ \hat{\Delta} = \hat{\hat{\Delta}} + \frac{\hat{\mu}}{n} . \]
So, since the predictive density is specified up to order $n^{-1}$, we can write
\[
\left( \frac{\nu}{n - \tau} \right)^{\frac{1}{2}} \Delta \sim \tau, \tag{9}
\]
a $t$ distribution with $\nu = (n - \tau)^2/n - 1$ degrees of freedom and where
\[
\tau = -\frac{4\nu + 3}{2} = p + 1 + \hat{\gamma}^i \hat{S}_\nu = p + 1 + da,
\]
where $S_\nu = (y_{n-i+1} - \mu) (y_{n-j+1} - \mu)$, $(\gamma^i) = (\gamma^j)^{-1}$, $(\gamma^i)$ is the variance–covariance matrix of the stationary distribution, $a = (a_k)_{1 \leq k \leq p}$ is the value of the ancillary statistics, that is
\[
a_k = (\hat{\gamma}^1)^{-1} (y_{n-j+1} - \hat{\mu})
\]
(see Appendix for more details) and
\[
\hat{\mu} = \frac{\hat{\sigma}}{1 - \sum \hat{\phi}_i} \sum \hat{\phi}_i \left( 1 - \sum \hat{\phi}_i \right) \sum \hat{\gamma}^i \left( \hat{\gamma}^1 \right)^{-1} (y_{n-j+1} - \hat{\mu}). \tag{11}
\]
We should mention that misspecifications in the model could produce intervals with the same error rate, for instance when the Gaussian assumption is violated.

### 4. Small sample simulation

We present the results of some simulations that show the behavior of the solutions for AR(1) and AR(2) processes. By using the S-PLUS Version 7, we have simulated samples of size $n = 18$, for different values of $\alpha$, $\mu$, $\phi_1$ for the AR(1) process. In all cases $\sigma = 1$, $y_0 = 0$, and the number of iterations is 10,000. For the AR(2) process we have simulated samples of size $n = 35$, for different values of $\alpha$, $\mu$, $\phi_1$ and $\phi_2$ for the AR(2) process. In all cases $\sigma = 1$, $y_0 = 0$, $y_{-1} = 0$ and the number of iterations is 10,000.

EST indicates the estimative pivot obtained by plugging in simply the estimation of the unknown parameters; BNC indicates the predictive pivot obtained here following the method suggested by Barndorff-Nielsen and Cox to obtain predictive densities. The numbers are the actual values of $\alpha$ one obtains by using the different pivots. The results are in Tables 1 and 2.

Therefore, the corrections are of the same order, independently of the value of the unknown parameters and the values of $\alpha$.

### Appendix

In this appendix we present in some detail the necessary calculations to obtain the prediction densities:

The mle of the autoregressive coefficients are the solution of the system
\[
\begin{align*}
\sum_{i=1}^{n} y_i y_{i-1} - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_{i-1} &= \sum_{j=1}^{p} \phi_j \left( \sum_{i=1}^{n} y_{i-j} y_{i-1} - \frac{1}{n} \sum_{i=1}^{n} y_{i-j} \sum_{i=1}^{n} y_{i-1} \right) \\
\sum_{i=1}^{n} y_i y_{i-2} - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_{i-2} &= \sum_{j=1}^{p} \phi_j \left( \sum_{i=1}^{n} y_{i-j} y_{i-2} - \frac{1}{n} \sum_{i=1}^{n} y_{i-j} \sum_{i=1}^{n} y_{i-2} \right) \\
& \vdots \\
\sum_{i=1}^{n} y_i y_{i-p} - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_{i-p} &= \sum_{j=1}^{p} \phi_j \left( \sum_{i=1}^{n} y_{i-j} y_{i-p} - \frac{1}{n} \sum_{i=1}^{n} y_{i-j} \sum_{i=1}^{n} y_{i-p} \right)
\end{align*}
\]
Table 2
AR(2) with $\mu = 0$

<table>
<thead>
<tr>
<th></th>
<th>$\phi_1$</th>
<th>EST</th>
<th>BNC</th>
<th>EST</th>
<th>BNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_2$</td>
<td>$\alpha = 0.95$</td>
<td>$\alpha = 0.90$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-0.75$</td>
<td>$-1.5$</td>
<td>0.9285</td>
<td>0.9484</td>
<td>0.8787</td>
<td>0.8976</td>
</tr>
<tr>
<td></td>
<td>$-0.75$</td>
<td>0.9356</td>
<td>0.9576</td>
<td>0.8855</td>
<td>0.9091</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>0.9338</td>
<td>0.9546</td>
<td>0.8823</td>
<td>0.9083</td>
</tr>
<tr>
<td></td>
<td>$1.5$</td>
<td>0.9336</td>
<td>0.9545</td>
<td>0.8838</td>
<td>0.9092</td>
</tr>
<tr>
<td>$-0.5$</td>
<td>$-1.25$</td>
<td>0.9248</td>
<td>0.9478</td>
<td>0.8703</td>
<td>0.8950</td>
</tr>
<tr>
<td></td>
<td>$-0.5$</td>
<td>0.9293</td>
<td>0.9528</td>
<td>0.8772</td>
<td>0.9032</td>
</tr>
<tr>
<td></td>
<td>$0.5$</td>
<td>0.9343</td>
<td>0.9565</td>
<td>0.8782</td>
<td>0.9069</td>
</tr>
<tr>
<td></td>
<td>$1.25$</td>
<td>0.9317</td>
<td>0.9546</td>
<td>0.8799</td>
<td>0.9067</td>
</tr>
<tr>
<td>$-0.25$</td>
<td>$-1$</td>
<td>0.9354</td>
<td>0.9559</td>
<td>0.8779</td>
<td>0.9055</td>
</tr>
<tr>
<td></td>
<td>$0$</td>
<td>0.9277</td>
<td>0.9539</td>
<td>0.8706</td>
<td>0.8990</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.9334</td>
<td>0.9561</td>
<td>0.8773</td>
<td>0.9039</td>
</tr>
<tr>
<td>$0$</td>
<td>$-0.75$</td>
<td>0.9313</td>
<td>0.9556</td>
<td>0.8833</td>
<td>0.9084</td>
</tr>
<tr>
<td></td>
<td>$0$</td>
<td>0.9327</td>
<td>0.9554</td>
<td>0.8792</td>
<td>0.9071</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>0.9326</td>
<td>0.9542</td>
<td>0.8821</td>
<td>0.9084</td>
</tr>
<tr>
<td>$0.25$</td>
<td>$-0.5$</td>
<td>0.9347</td>
<td>0.9555</td>
<td>0.8815</td>
<td>0.9057</td>
</tr>
<tr>
<td></td>
<td>$0$</td>
<td>0.9360</td>
<td>0.9580</td>
<td>0.8818</td>
<td>0.9093</td>
</tr>
<tr>
<td></td>
<td>$0.5$</td>
<td>0.9339</td>
<td>0.9550</td>
<td>0.8822</td>
<td>0.9094</td>
</tr>
<tr>
<td>$0.5$</td>
<td>$-0.25$</td>
<td>0.9301</td>
<td>0.9541</td>
<td>0.8822</td>
<td>0.9065</td>
</tr>
<tr>
<td></td>
<td>$0$</td>
<td>0.9281</td>
<td>0.9528</td>
<td>0.8770</td>
<td>0.9014</td>
</tr>
<tr>
<td></td>
<td>$0.25$</td>
<td>0.9329</td>
<td>0.9569</td>
<td>0.8824</td>
<td>0.9068</td>
</tr>
<tr>
<td>$0.75$</td>
<td>$0$</td>
<td>0.9325</td>
<td>0.9548</td>
<td>0.8767</td>
<td>0.9048</td>
</tr>
</tbody>
</table>

and $\hat{\mu}$ and $\hat{\sigma}^2$ are

$$
\hat{\mu} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{p} \phi_j y_{i-j}}{n(1 - \sum \phi_j)},
$$

$$
\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \left[ y_j - \hat{\mu} - \sum_{j=1}^{p} \phi_j (y_{i-j} - \hat{\mu}) \right]^2.
$$

To determine a minimal predictive sufficient statistics we use the Theorem 1. Then we consider the density of $Y = (Y_1, Y_2, \ldots, Y_n)$ conditioned to $Y_{n+1} = y_{n+1}$, and we have a $(p + 1)^2, p + 2$ exponential model, see Barndorff-Nielsen and Cox (1994), page 65, for the definition, with

$$
U(y) = \left( y_n, y_{n-1}, \ldots, y_{n-p+1}, \sum_{j=1}^{n} y_j, \sum_{j=1}^{n} y_{i-j}, \sum_{j=1}^{n} y_{i-j} y_{i-k} \right), j = 1, 2, \ldots, p; k = 1, 2, \ldots, p
$$

as minimal sufficient statistics. So,

$$(\hat{\mu}, \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}^2, y_n, y_{n-1}, \ldots, y_{n-p+1})$$

is a $(2p + 1)$ predictive sufficient statistics, $(y_n, y_{n-1}, \ldots, y_{n-p+1})$ being a minimal transitive statistics. To make the decomposition $(\hat{\theta}, A)$, $A$ being a vector of ancillary statistics, we can take

$$
A = \Gamma^{-1} (Y_{(p)} - \hat{\mu} I_p), \quad Y'_{(p)} = (Y_n, Y_{n-1}, \ldots, Y_{n-p}),
$$

(12)

where $\Gamma := (\Gamma_0)$ is the asymptotic covariance matrix of $Y_{(p)}$, that is the covariance of the stationary distribution, and the prime denotes transposed. The distribution of $A$ is approximately $N(0, I_p)$, so $A$ is approximately ancillary.

The asymptotic information matrix, where we indicate only the non-null terms, is given by

$$
(i') = \begin{pmatrix}
\frac{\sigma^2}{n(1 - \sum \phi_j)^2} & \frac{\sigma^2 (\Gamma^m)}{n} & \frac{\sigma^2}{2n} \\
\frac{\sigma^2 (\Gamma^m)}{n} & \frac{\sigma^2}{2n} & \frac{\sigma^2}{2n}
\end{pmatrix},
$$
where \((F^n) = F^{-1}\). After rather tedious calculations, we obtain that
\[
\begin{align*}
I_{\mu, \theta} & = I_{\mu, \theta} = I_{\mu, \hat{\theta}} = 0, \\
I_{\phi, \theta} & = 2n \left( 1 - \sum \phi \right) / \sigma^2; \\
I_{\phi, \hat{\phi}} & = 0, \\
I_{\phi, \hat{\theta}} & = 2n \left( 1 - \sum \phi \right)^2 / \sigma^3.
\end{align*}
\]

The symbol \(\hat{=}\) indicates that we consider the leading term of the above quantities as powers of \(n\) by neglecting terms that contributes to \(Q\) given in (7), with \(o(n^{-1})\). Moreover, the above derivatives are calculated in \(\hat{\theta} = \theta\). The expression of \(Q\) (by formula (7)) is given by
\[
Q = \frac{1}{2} \left\{ \mu \sigma H_{\mu, \mu} + r^3 \left( H_{\phi, \phi} - H_{\phi} \left( I_{\phi, \hat{\mu}, \mu} \mu + I_{\phi, \hat{\phi}, \phi} \phi \right) \right) + r^3 \left( H_{\sigma, \sigma} - H_{\sigma} \left( I_{\sigma, \hat{\mu}, \mu} \mu + I_{\sigma, \hat{\phi}, \phi} \phi + I_{\sigma, \hat{\sigma}, \sigma} \sigma \right) \right) \right\}
\]
with (using formulas (8))
\[
H_{\mu, \mu} = \frac{g}{\sigma} \left( 1 - \sum_{j=1}^{p} \phi \right)^2 \Delta,
\]
where \(\Delta = \left( y_{n+1} - \mu - \sum_{j=1}^{p} \phi y_{n-j+1} - \mu \right) / \sigma\),
\[
H_{\phi, \phi} = g \left\{ -\frac{\Delta}{\sigma} S + \Delta \left( y_{n-i+1} + \Delta \right) \right\},
\]
where \(S = (y_{n-i+1} - \mu) (y_{n-j+1} - \mu)\),
\[
H_{\phi} = g \left( y_{n-i+1} - \mu \right)
\]
and finally
\[
H_{\sigma, \sigma} = -\frac{g}{\sigma} \Delta^3
\]
\[
H_{\sigma} = g \Delta.
\]

Then
\[
r = \frac{Q}{g} = \frac{\sigma}{n} \left( u + v \Delta - \frac{\Delta^3}{4} \right),
\]
with
\[
\begin{align*}
v & = \frac{\sigma}{1 - \sum \phi} \left( \left( 1 - \sum \phi \right) \left( \sum_k \left( \Gamma^{-\frac{1}{2}} \right)^i \right) \left( \Gamma^{-\frac{1}{2}} \right)^j \left( y_{n-j+1} - \mu \right) \right), \\
v & = -\frac{5}{4} \left( \frac{p}{2} - \frac{r_{ij}}{2} \right).
\end{align*}
\]

References