

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Helske, Jouni; Nyblom, Jukka

Title: Improved Frequentist Prediction Intervals for Autoregressive Models by Simulation

Year: 2015

Version:

Please cite the original version:

Helske, J., & Nyblom, J. (2015). Improved Frequentist Prediction Intervals for Autoregressive Models by Simulation. In S. J. Koopman, & N. Shephard (Eds.), *Unobserved Components and Time Series Econometrics* (pp. 291-309). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199683666.003.0013>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Improved Frequentist Prediction Intervals for Autoregressive Models by Simulation

Jouni Helske* and Jukka Nyblom
Department of Mathematics and Statistics
University of Jyväskylä

Abstract

It is well known that the so called plug-in prediction intervals for autoregressive processes, with Gaussian disturbances, are too narrow, i.e. the coverage probabilities fall below the nominal ones. However, simulation experiments show that the formulas borrowed from the ordinary linear regression theory yield one-step prediction intervals, which have coverage probabilities very close to what is claimed. From a Bayesian point of view the resulting intervals are posterior predictive intervals when uniform priors are assumed for both autoregressive coefficients and logarithm of the disturbance variance. This finding opens the path how to treat multi-step prediction intervals which are obtained easily by simulation either directly from the posterior distribution or using importance sampling. A notable improvement is gained in frequentist coverage probabilities. An application of the method to forecasting the annual gross domestic product growth in the United Kingdom and Spain is given for the period 2002–2011 using the estimation period 1962–2001.

Key words: Importance sampling; Jeffreys's prior distribution; Predictive distribution; Multi-step predictions; Uniform prior distribution.

*Corresponding author: Jouni Helske, jouni.helske@jyu.fi

1 Introduction

A traditional approach to time series forecasting usually involves a selection of a family of suitable models, e.g. the class of autoregressive integrated moving average (ARIMA) models. Then using different model selection criteria within the family based of autocorrelation and partial autocorrelation functions together with formal criteria such as Akaike or Bayesian information criterion, the analyst chooses a suitable “best fitting” representative from the model family, estimates the parameters, makes diagnostic checks, and if he is happy with his choice computes predictions and the prediction intervals. The prediction intervals are usually computed as if the chosen model were correct and the parameters completely known, with no reference to the process of model selection. Chatfield (1993, 1996) has strongly criticized the omission of model uncertainty in forecasting. Clements and Hendry (1999, sections 1.3.8 and 2.2) introduce a detailed taxonomy of forecast errors, and stress the effects of the structural breaks in time series forecasting. As a remedy they propose robustifying forecasts for example by differencing and intercept corrections.

It is the common view of references given in the previous paragraph that the parameter uncertainty is often a minor source of prediction errors in practical applications when the sample size is large enough. Clements and Hendry (1999, p. 128) remark that although the parameter uncertainty is unlikely to lead to serious forecast failure it may have larger effect in conjunction with model misspecification. Nevertheless, we believe that it is justified to handle also this part of the model uncertainty. In textbooks it is a common topic, see for example Harvey (1993, p. 58-59). Here we show how to make corrections in a fairly simple way under autoregressive (AR) models.

Several proposals have been made for improving prediction intervals when parameters are estimated. One group of solutions focus on finding a more accurate prediction mean squared error in the presence of estimation; see, for example Phillips (1979), Fuller and Hasza (1981), Ansley and Kohn (1986), Quenneville and Singh (2000), and Pfeiffermann and Tiller (2005). Both analytic and bootstrap approaches are tried.

Barndorff-Nielsen and Cox (1996) give general results for prediction intervals in the presence of estimated parameters. These results are further developed for time series models by Vidoni (2004, 2009). Unfortunately fairly complicated expressions appear already in rather simple models. Bootstrap solutions are given by several authors; see for example Beran (1990),

Masarotto (1990), Grigoletto (1998), Kim (2004), Pascual, Romo, and Ruiz (2004), Clements and Kim (2007), Kabaila and Syuhada (2008), and Rodriguez and Ruiz (2009).

Our aim is to find solutions for prediction interval problems using a Bayesian viewpoint with a requirement of objectivity for our predictions. Berger (2006) strongly encourages the use of the term “objective Bayes” in such situations. Operationally “objectivity” in our application means adopting priors which produce prediction intervals which exhibit approximately same coverage probabilities in both Bayesian and frequentist sense.

There is a vast literature on matching posterior and frequentist inferences to some degree of approximation. Methods based on invariance arguments, on information criteria or divergence measures as well as on asymptotic expansions are tried. The introduction section of Fraser et al. (2010) gives a short review on these as well as a list of relevant references to their own works and others. The starting point of their approach is to replace p value computations in a sample space with with posterior integration over the parameter space. Matching the Bayesian and frequentist coverage probabilities of prediction intervals in regression models with independent cases are treated in Datta and Mukerjee (2003) and Datta and Mukerjee (2004). A common feature in all these approaches is that often the so called Jeffreys’s prior or its modification shows up as a solution. As we will see it happens also in this article. In a recent article of Arellano and Bonhomme (2009), where the main issue is the bias reduction in panel data models, the authors use a prior related to Jeffreys’s prior. A predecessor in the bias reduction with the help of Jeffreys’s prior is given by Firth (1993).

Early examples using the Bayesian approach in autoregressive models are given by Zellner (1971, p. 188) and Chow (1974). Later Broemeling and Land (1984) showed, among other things, that using a normal-gamma prior for the parameters the one-step ahead prediction follows a t distribution (understood as a location-scale family). They further deduced that all predictions up to k step ahead have a joint predictive density which consists of the product of k univariate t densities. Thompson and Miller (1986) propose a simulation method for computing prediction intervals based on the Bayesian arguments called “sampling the future”. This differs from our approach. We compute the prediction interval by simulating directly the posterior prediction probability which is more accurate and considerably less time consuming. Liu (1994) develops an approximation to their method, and Snyder, Ord, and Koehler (2001) make an approximate extension of it to ARIMA mod-

els. They also compare several types of prediction intervals in terms of the frequentist coverage probabilities.

After the emergence of Markov Chain Monte Carlo simulation the Bayesian paradigm has gained increasing popularity in time series and econometrics; see for example Geweke (2005) and Prado and West (2010) and references therein. A thorough exposition of the Bayesian forecasting is given by Geweke and Whiteman (2006). Nevertheless, in forecasting under AR model either direct simulation or importance sampling is fast and sufficient. Based on independent simulation replicates it also renders rather simple formulas for assessing the Monte Carlo simulation error.

2 Motivation

Consider an AR(p) process

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t, \quad t = 1, \dots, n, \dots, \quad (2.1)$$

where the errors ε_t are independently drawn from $N(0, \sigma^2)$, and coefficients $\beta_j, j = 0, \dots, p$, are arbitrary fixed values. Assume that we have observed y_{-p+1}, \dots, y_n . Write $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})'$, and let \mathbf{X} be the matrix with rows $\mathbf{x}_t', t = 1, \dots, n$. Further let $\mathbf{Y} = (y_1, \dots, y_n)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$. Then the model (2.1) can be written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. If we condition on the starting values $\mathbf{y}'_0 = (y_{-p+1}, \dots, y_0)$ we are led to the conditional likelihood

$$L_c(\boldsymbol{\beta}, \sigma) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right). \quad (2.2)$$

The conditional maximum likelihood estimates for $\boldsymbol{\beta}$ coincide with the least squares estimates

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (2.3)$$

and the maximum likelihood estimate for σ^2 , corrected by the degrees of freedom, is

$$s^2 = \frac{(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})}{n - p - 1}. \quad (2.4)$$

The predictive value for y_{n+1} is $\mathbf{x}'_{n+1}\widehat{\boldsymbol{\beta}}$, and the standard prediction interval with approximate coverage probability $1 - 2\alpha$ is

$$\mathbf{x}'_{n+1}\widehat{\boldsymbol{\beta}} \pm sz_\alpha, \quad (2.5)$$

where z_α is the α quantile of the standard normal distribution. In practice, the true coverage may be considerably below the nominal value $1 - 2\alpha$.

Let us suppose for a moment that we have an ordinary regression model with some truly exogeneous variables. Then using the same notation, \mathbf{X} and \mathbf{x}_{n-1} , as before the exact coverage probability is obtained by

$$\mathbf{x}'_{n+1}\widehat{\boldsymbol{\beta}} \pm st_{\alpha, n-p-1}\sqrt{1 + \mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1}}, \quad (2.6)$$

As is well known, the extra factor here compared to (2.5) involving the exogeneous variable takes into account the estimation error in regression coefficients. In addition, a minor correction is done by replacing the normal quantile with that from Student's t with $n - p - 1$ degrees of freedom. Although the assumptions of AR models do not satisfy the assumptions leading to (2.6), our simulations show that the very same intervals (2.6) have practically correct coverage probabilities also under the AR models.

Combining the formulas (2.2)–(2.4) with the identity

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \\ &\quad + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{aligned}$$

the conditional likelihood function (2.2) can be written as

$$\begin{aligned} L_c(\boldsymbol{\beta}, \sigma) &= (2\pi)^{-\frac{n}{2}}\sigma^{-n} \exp\left(-\frac{(n-p-1)s^2}{2\sigma^2}\right) \\ &\quad \times \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right). \end{aligned} \quad (2.7)$$

Until now we have had the frequentist approach. But it is also illuminating to see the prediction interval (2.6) from a Bayesian point of view, where $\boldsymbol{\beta}$ and σ are now treated as a random vector and variable. If we multiply the likelihood $L_c(\boldsymbol{\beta}, \sigma)$ by the improper prior $p(\boldsymbol{\beta}, \sigma) = 1/\sigma$, we find, as is well known in the ordinary regression, that *a posteriori*

$$\begin{aligned} \frac{(n-p-1)s^2}{\sigma^2} \Big| \mathbf{Y} &\sim \chi^2(n-p-1), \\ \boldsymbol{\beta} | \mathbf{Y}, \sigma &\sim N(\widehat{\boldsymbol{\beta}}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \end{aligned} \quad (2.8)$$

These together lead to the result that the interval (2.6) is a Bayesian predictive interval with exact coverage probability $1 - 2\alpha$. The corresponding

frequentist coverage probability is, however, only approximate

$$P\left(y_{n+1} \in \mathbf{x}'_{n+1}\widehat{\boldsymbol{\beta}} \pm st_{\alpha, n-p-1}\sqrt{1 + \mathbf{x}'_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{n+1}} \mid \boldsymbol{\beta}, \sigma\right) \approx 1 - 2\alpha. \quad (2.9)$$

However, in section 5 we will find that the approximation (2.9) is very good when the coefficients are well within the stationarity region. The approximation seems to be worst in the nearly unit root cases. The explanation is likely to be as follows. If the coefficients are not too close to the boundary of the stationarity region, the sampling distribution of $\widehat{\boldsymbol{\beta}}$ is approximately as in (2.8) when the roles of $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$ are interchanged. On the other hand this not true in the nearly unit root case. So the question arises what is “good” prior in AR models in general. It turns out that using a certain modification of Jeffreys’s prior seems to be preferred to uniform prior in nearly unit root models. One solution to the nearly unit root problem is provided by Clements and Hendry (1999, p. 92) who recommend in some cases to impose a unit root albeit it were not warranted by the unit root test.

At this point it might be interesting to recall the vigorous debate that broke out in the early 1990’s between Peter Phillips and Christopher Sims on the value of unit root econometrics as such and its relation to Bayesian approach in statistics and econometrics, see Sims and Uhlig (1991), Phillips (1991), Sims (1991) and their references. The unit root inference is not an issue here. The Bayesian controversy focused on the choice of an appropriate prior distribution. Nevertheless, we do not want to revive this controversy here simply because we are not so much interested in inferences on autoregressive coefficients themselves but rather in prediction intervals. The role of the prior in this article is to produce a good weighting scheme for the predictive distribution.

3 Predictive distributions

3.1 Uniform prior

In this section we develop prediction formulas under the Bayesian paradigm employing the noninformative uniform prior for $(\boldsymbol{\beta}, \log \sigma)$, i.e. the prior takes the form $p(\boldsymbol{\beta}, \sigma) = 1/\sigma$, for $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ and $\sigma > 0$. We have already seen that when $k = 1$ the predictive distribution is Student’s t . But for $k > 1$

the predictive distribution is not known to be any common distribution. Therefore we have to rely on simulation.

We use a generic notation $p(\cdot)$ for a density. The Bayesian predictive density of y_{n+k} is defined as the conditional density of y_{n+k} given \mathbf{Y} . It is obtained by integration as

$$p(y_{n+k} | \mathbf{Y}) = \int p(y_{n+k} | \boldsymbol{\beta}, \sigma, \mathbf{Y}) p(\boldsymbol{\beta}, \sigma | \mathbf{Y}) d\boldsymbol{\beta} d\sigma, \quad (3.1)$$

where $p(\boldsymbol{\beta}, \sigma | \mathbf{Y})$ is the posterior density of $(\boldsymbol{\beta}, \sigma)$. In AR models with normally distributed errors the density $p(y_{n+k} | \boldsymbol{\beta}, \sigma, \mathbf{Y})$ can be written explicitly. Recall the prediction formulas given $\boldsymbol{\beta}, \sigma^2, \mathbf{Y}$ (see for example Box et al. (2008, chap. 5)) and designate $\hat{y}_n(k) = E(y_{n+k} | \mathbf{Y}, \boldsymbol{\beta})$, $k = 0, \pm 1, \pm 2, \dots$. Note that when $k \leq 0$ we have $\hat{y}_n(k) = y_{n+k}$. The equation (2.1) immediately yields recursion for predicted values

$$\hat{y}_n(k) = \beta_0 + \beta_1 \hat{y}_n(k-1) + \dots + \beta_p \hat{y}_n(k-p), \quad k = 1, 2, \dots$$

It is also plain from (2.1) that for some constants ψ_1, ψ_2, \dots depending on $\boldsymbol{\beta}$ we have

$$y_{n+k} - \hat{y}_n(k) = \varepsilon_{n+k} + \psi_1 \varepsilon_{n+k-1} + \dots + \psi_{k-1} \varepsilon_{n+1}.$$

Box et al. (2008) gives recursions for the coefficients ψ_j as follows

$$\psi_j = \beta_1 \psi_{j-1} + \dots + \beta_p \psi_{j-p}, \quad j \geq 1, \quad \psi_0 = 1, \quad \text{and } \psi_j = 0 \text{ for } j < 0.$$

We find $\psi_1 = \beta_1, \psi_2 = \beta_1^2 + \beta_2$ and so on. The prediction error variance, given $\boldsymbol{\beta}, \sigma^2$, is then $\sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2) = \sigma^2 v_k^2(\boldsymbol{\beta})$.

In the further development it is useful to introduce a more detailed notation $\hat{y}_n(k; \boldsymbol{\beta}) = E(y_{n+k} | \mathbf{Y}, \boldsymbol{\beta})$. Then

$$y_{n+k} | \boldsymbol{\beta}, \sigma^2, \mathbf{Y} \sim N(\hat{y}_n(k; \boldsymbol{\beta}), \sigma^2 v_k^2(\boldsymbol{\beta})).$$

Combining this with (3.1) and changing the order of integration we find

$$P(y_{n+k} \leq b | \mathbf{Y}) = E \left[\Phi \left(\frac{b - \hat{y}_n(k; \boldsymbol{\beta})}{\sigma v_k(\boldsymbol{\beta})} \right) \middle| \mathbf{Y} \right], \quad (3.2)$$

where Φ is the cumulative distribution function of the standard normal distribution. When $k > 1$ we cannot do the integration involved analytically. Nevertheless, a Monte Carlo solution is straightforward. We can proceed as follows:

1. Draw independent q_i , $i = 1, \dots, N$ from $\chi^2(n - p - 1)$, and let $\sigma_i^2 = (n - p - 1)s^2/q_i$.
2. Draw β_i from $N(\hat{\beta}, \sigma_i^2(\mathbf{X}'\mathbf{X})^{-1})$, independently for $i = 1, \dots, N$.
3. Compute the average

$$\bar{P}_N(b) = \frac{1}{N} \sum_{i=1}^N \Phi \left(\frac{b - \hat{y}_n(k; \beta_i)}{\sigma_i v_k(\beta_i)} \right). \quad (3.3)$$

The prediction interval is then found by solving separately both $\bar{P}_N(b) = \alpha$ and $\bar{P}_N(b) = 1 - \alpha$. Let the solutions be $\hat{b}_\alpha, \hat{b}_{1-\alpha}$ respectively. Then $(\hat{b}_\alpha, \hat{b}_{1-\alpha})$ is the prediction interval with posterior coverage probability $1 - 2\alpha$ when N is large. Broemeling and Land (1984) noted that multi-step predictions can be constructed through a sequence of one-step predictions each step involving t distribution. However, they do not suggest any computational method how to utilize this in practice. Thompson and Miller (1986) continued the work of Broemeling and Land by proposing the Bayesian simulation of the future values y_{n+1}, \dots, y_{n+k} . The prediction limits are then derived from the quantiles of the simulated values. This is more time consuming as well as being less accurate than what we suggest here.

3.2 Prediction With General Prior

Combine the conditional likelihood (2.2) and a general prior $p(\beta, \sigma)/\sigma$. Then by (2.7) and (2.8) the joint posterior is

$$\begin{aligned} p(\beta, \sigma | \mathbf{Y}) &\propto \sigma^{-(n-p)} e^{-(n-p-1)s^2/(2\sigma^2)} \\ &\quad \times \sigma^{-p-1} \exp \left(-\frac{1}{2\sigma^2} (\hat{\beta} - \beta)' (\mathbf{X}'\mathbf{X}) (\hat{\beta} - \beta) \right) \\ &\quad \times p(\beta, \sigma). \end{aligned}$$

It can be evaluated using importance sampling. In the simulation algorithm only the item 3 need be changed to

- 3' Compute the weighted average

$$\bar{P}_{N,w}(b) = \frac{\sum_{i=1}^N w_i \Phi \left(\frac{b - \hat{y}_n(k; \beta_i)}{\sigma_i v_k(\beta_i)} \right)}{\sum_{i=1}^N w_i}, \quad (3.4)$$

$$w_i = p(\beta_i, \sigma_i). \quad (3.5)$$

The prediction interval is then solved as before.

It is also possible to incorporate the starting values $(y_{-p+1}, \dots, y_0) = \mathbf{y}'_0$ into the likelihood. Denote $\text{cov}(\mathbf{y}_0 | \boldsymbol{\beta}, \sigma) = \sigma^2 \mathbf{V} = \sigma^2 \mathbf{V}(\boldsymbol{\beta})$, and $E(y_i | \boldsymbol{\beta}) = \mu = \mu(\boldsymbol{\beta})$, $i = -p + 1, \dots, 0$. The matrix \mathbf{V} can be obtained from the Yule-Walker equations, see Box et al. (2008, p. 58), and $\mu = \beta_0 / (1 - \beta_1 - \dots - \beta_p)$. To obtain the full likelihood, the conditional likelihood (2.2) is just multiplied by

$$p(\mathbf{y}_0 | \boldsymbol{\beta}, \sigma) = (2\pi)^{-\frac{p}{2}} \sigma^{-p} |\mathbf{V}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (\mathbf{y}_0 - \mu \mathbf{1})' \mathbf{V}^{-1} (\mathbf{y}_0 - \mu \mathbf{1}) \right).$$

This leads to changing the weights in step 3'. The new weights are

$$w_i = I(\boldsymbol{\beta}_i \in \mathbb{R} \times S_p) p(\boldsymbol{\beta}_i, \sigma_i) p(\mathbf{y}_0 | \boldsymbol{\beta}_i, \sigma_i) \quad (3.6)$$

where $I(\cdot)$ is an indicator and S_p is the stationarity region of AR(p) process. The first coordinate of $\boldsymbol{\beta}_i$ can, of course, take any real value.

3.3 Standard Errors for Monte Carlo Prediction Limits

Here we give simple formulas for approximate standard errors when computing the prediction limits by Monte Carlo simulation. Let b_α be such that $P(y_{n+k} \leq b_\alpha | \mathbf{Y}) = \alpha$ in (3.2), and \hat{b}_α be such that $\bar{P}_N(\hat{b}_\alpha) = \alpha$ in (3.3). The first order Taylor expansion at \hat{b}_α leads to

$$\hat{b}_\alpha - b_\alpha \approx \frac{\alpha - \bar{P}_N(b_\alpha)}{\bar{P}'_N(\hat{b}_\alpha)},$$

where \bar{P}'_N is the derivative of \bar{P}_N . The variance of the numerator on the right side can be estimated from the sample values by S^2/N where

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N \left[\Phi \left(\frac{\hat{b}_\alpha - \hat{y}_n(k; \boldsymbol{\beta}_i)}{\sigma_i v_k(\boldsymbol{\beta}_i)} \right) - \alpha \right]^2.$$

Define

$$\text{s.e.}(\hat{b}_\alpha) = \frac{S/\sqrt{N}}{\bar{P}'_N(\hat{b}_\alpha)} = \frac{S}{\sum_{i=1}^N \frac{1}{\sigma_i v_k(\boldsymbol{\beta}_i)} \varphi \left(\frac{\hat{b}_\alpha - \hat{y}_n(k; \boldsymbol{\beta}_i)}{\sigma_i v_k(\boldsymbol{\beta}_i)} \right) / \sqrt{N}},$$

where φ is the density of the standard normal distribution. Omitting the details we can show that $(\hat{b}_\alpha - b_\alpha)/\text{s.e.}(\hat{b}_\alpha)$ tends to the standard normal distribution as $N \rightarrow \infty$.

In case we use a weighted average as in (3.5) and (3.6), a similar technique leads to the standard error

$$\begin{aligned} \text{s.e.}(\hat{b}_\alpha) &= \frac{S}{\sum_{i=1}^N \frac{w_i}{\sigma_i v_k(\boldsymbol{\beta}_i)} \varphi\left(\frac{\hat{b}_\alpha - \hat{y}_n(k; \boldsymbol{\beta}_i)}{\sigma_i v_k(\boldsymbol{\beta}_i)}\right) / \sqrt{N}}, \\ S^2 &= \frac{1}{N-1} \sum_{i=1}^N \left[\alpha w_i - w_i \Phi\left(\frac{\hat{b}_\alpha - \hat{y}_n(k; \boldsymbol{\beta}_i)}{\sigma_i v_k(\boldsymbol{\beta}_i)}\right) \right]^2. \end{aligned}$$

4 Priors

Here we give some examples of the priors that are likely to give improvements for frequentist coverage probabilities. A popular principle to generate priors, often improper, is to follow Jeffreys's rule which leads to the square root of the determinant of the information matrix. Applying this to the conditional likelihood (2.2) we first find that

$$\frac{\partial^2 L_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\frac{1}{\sigma^2} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t'.$$

The information matrix is obtained by taking the expectation given the parameters and changing the sign. Assuming stationarity yields

$$-E \left[\frac{\partial^2 L_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right] = \frac{n}{\sigma^2} E[\mathbf{x}_1 \mathbf{x}_1'] = \frac{n}{\sigma^2} \begin{pmatrix} 1 & \mu \mathbf{1}' \\ \mu \mathbf{1} & \mathbf{V} + \mu^2 \mathbf{1} \mathbf{1}' \end{pmatrix}.$$

The determinant of the matrix involved is easily seen to equal $|\mathbf{V}|$ (note that \mathbf{V} depends only on β_j , $j = 1, \dots, p$). Now we take the convention that the parameter groups $\{\sigma\}$, $\{\beta_0\}$ and $\{\beta_1, \dots, \beta_p\}$ are independent *a priori*, and that $\log \sigma$, β_0 are uniform. This leads to the prior we here call Jeffreys's prior.

$$p_J(\boldsymbol{\beta}, \sigma) / \sigma = I(\sigma > 0) I(\boldsymbol{\beta} \in \mathbb{R} \times S_p) \sigma^{-1} \sqrt{|\mathbf{V}(\boldsymbol{\beta})|}. \quad (4.1)$$

The uniform prior over the stationarity region is

$$p_U(\boldsymbol{\beta}, \sigma) / \sigma = I(\sigma > 0) I(\boldsymbol{\beta} \in \mathbb{R} \times S_p) \sigma^{-1}.$$

In our simulations we combine these priors with the full likelihood, i.e. the weights are $p(\mathbf{y}_0 | \boldsymbol{\beta}_i, \sigma_i)p_J(\boldsymbol{\beta}_i, \sigma_i)$ and $p(\mathbf{y}_0 | \boldsymbol{\beta}_i, \sigma_i)p_U(\boldsymbol{\beta}_i, \sigma_i)$ for Jeffreys's and uniform on the stationarity region priors, respectively. Note that when Jeffreys's prior is used the determinant $|\mathbf{V}|$ cancels when forming the weights.

For the stationary AR(1) model Jeffreys's marginal prior density of β_1 is

$$\begin{aligned} p_J(\beta_1) &= \frac{1}{\pi\sqrt{1-\beta_1^2}}, & |\beta_1| < 1 \\ &= 0, & \text{otherwise,} \end{aligned}$$

and for AR(2) the marginal prior density of (β_1, β_2) is

$$\begin{aligned} p_J(\beta_1, \beta_2) &= \frac{1}{(1+\beta_2)\sqrt{(1-\beta_2)^2-\beta_1^2}}, & \beta_1 + \beta_2 < 1, \beta_2 - \beta_1 < 1, \\ & & |\beta_2| < 1, \\ &= 0, & \text{otherwise.} \end{aligned}$$

Note that the first density is proper whereas the second one is not; the latter property is true for all AR(p) with $p > 1$.

Without a stationarity restriction Berger and Yang (1994) defined an alternative proper marginal prior density for the AR(1) model as

$$\begin{aligned} p_R(\beta_1) &= \frac{1}{2\pi\sqrt{1-\beta_1^2}}, & |\beta_1| < 1, \\ &= \frac{1}{2\pi|\beta_1|\sqrt{\beta_1^2-1}}, & |\beta_1| > 1, \end{aligned} \tag{4.2}$$

which they call a reference prior. It is easily seen that the reference prior is invariant under the transformation $\beta_1 \rightarrow 1/\beta_1$. Note that the reference prior should be combined with the conditional likelihood (2.2), and then it produces peaks also in the posterior at $\beta_1 = \pm 1$. There is no use to combine it with the full likelihood because then it reduces to Jeffreys's prior. There seems to be no generalization available of the reference prior to higher order models.

5 Simulation experiments

In this section we compare different priors as regards their ability to match the frequentist coverage probabilities to the Bayesian ones. Thus, we now

turn back to the frequentist interpretation of the probability. Suppose that we have a realization from an autoregressive process with parameters β, σ which are fixed but unknown. Despite of this we wish to compute the prediction interval $(b_\alpha, b_{1-\alpha})$ via a Bayesian route from the probability distribution (3.2). Although the posterior coverage probability is exactly $1 - 2\alpha$, the corresponding frequentist probability usually is not. It refers to infinite sequence of new realizations from the same model with parameters β, σ , and it is defined by the probability

$$P(b_\alpha \leq y_{n+k} \leq b_{1-\alpha} \mid \beta, \sigma),$$

where all $y_{n+k}, b_\alpha, b_{1-\alpha}$ are random. In short, the frequentist coverage probability is an average coverage probability over the realizations. The conditional frequentist coverage probability $P(b_\alpha \leq y_{n+k} \leq b_{1-\alpha} \mid \mathbf{Y}, \beta, \sigma)$ is a random variable, and in an actual application we do not know this probability. In our simulation experiment the values $b_\alpha, b_{1-\alpha}$ are replaced by $\hat{b}_\alpha, \hat{b}_{1-\alpha}$ obtained from (3.3) (apart from the case of the uniform prior with $k = 1$).

The chosen priors are uniform, uniform on the stationarity region, Jeffreys's prior (4.1) and the reference prior (4.2) of Berger and Yang (1994) for AR(1). In all comparisons we have used 50,000 replicates, where each replicate is a realization from a stationary AR(p) process of length $n + p$. Within each replicate the Monte Carlo sample size is $N = 50$. In this type of simulation experiments N need not be large, because the main variation in coverage probabilities is due to the different replicates. Nevertheless, in an actual application N should be considerably larger as is seen in the next section. All computations are done in the R environment (R Development Core Team, 2012).

Start with AR(1). Figure 5.1 shows the coverage probabilities of one-step ahead prediction intervals (2.6) based on the t distribution for several values of β_1 , when $n = 30$. Here, and in all other experiments, we are aiming for the coverage probability of 0.9. For negative β_1 the coverage probabilities tend to be slightly above the nominal value whereas when β_1 is close to 1, they drop below the nominal value. However, note that even for $\beta_1 = 0.9$ the coverage probability is 0.894. The standard errors of coverage probabilities are less than 3×10^{-4} in all cases. Although not shown in the figure the prediction intervals appear to have approximately equal tail probabilities. The coverage probabilities of the standard prediction intervals stay below 0.88.

Figure 5.2 shows the coverage probabilities of multi-step prediction for AR(1) processes with different prior choices and four different parameter

values for β_1 . The standard error of coverage probability was less than 7×10^{-4} in all cases. The reference prior seems to be slightly preferred compared to others. Especially, when β_1 is near to 1, the reference prior leads to coverage probabilities that are closest to the nominal frequentist coverage probability. Elsewhere the priors give almost the same coverage probabilities.

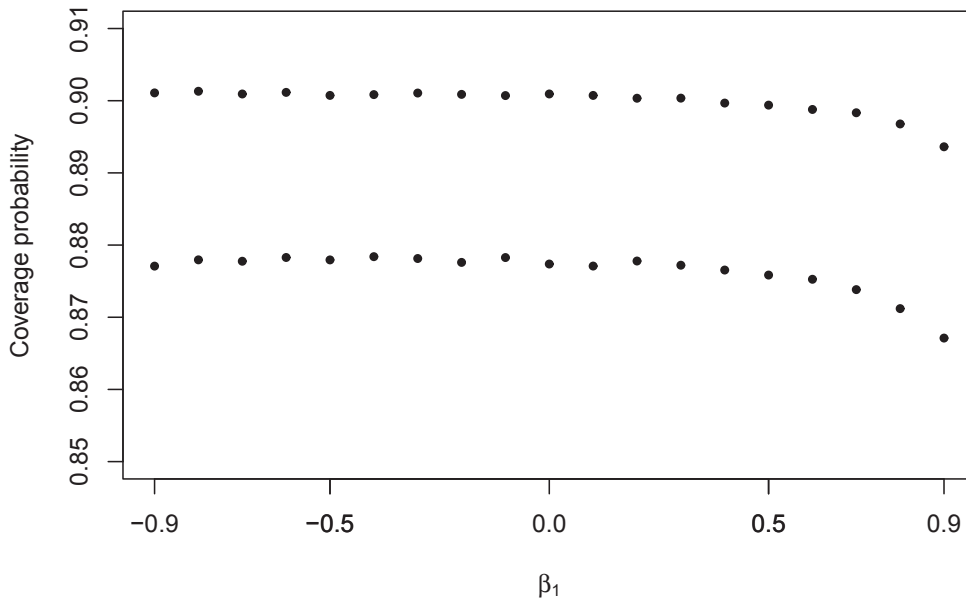


Figure 5.1: Coverage probabilities of one-step ahead predictions in AR(1) models when $n = 30$. The nominal coverage is 0.9. Dots are the estimated values, based on 50000 replicates. The upper dots relate to the Bayesian intervals with uniform prior and the lower ones to standard intervals.

In order to examine the differences between plug-in method and the Bayesian method with different priors in the case of AR(2) processes, we use nine different parameter combinations for β_1 and β_2 . The parameters are defined through the roots of the characteristic equation

$$\beta(r) = 1 - \beta_1 r - \beta_2 r^2 = 0.$$

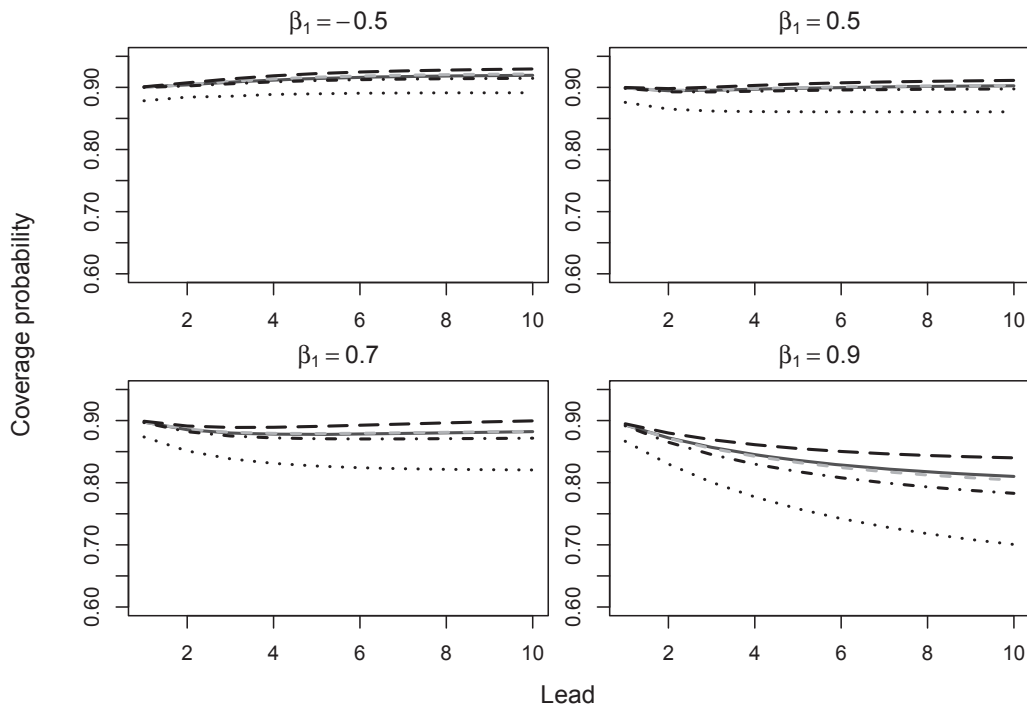


Figure 5.2: Coverage probabilities of multi-step ahead predictions in AR(1) models with different values of β_1 when $n = 30$. The nominal coverage is 0.9. The top dashed black line is based on the reference prior and the bottom dotted line represents the standard coverage probabilities. The lines based on the uniform (solid dark grey line), Jeffreys (light grey dashed line) and uniform stationary (black dot-and-dash line) prior are almost indiscernible apart from the case $\beta_1 = 0.9$.

Let the roots be r_1 and r_2 . Then the parameters β_1 and β_2 can be written as

$$\beta_1 = \frac{1}{r_1} + \frac{1}{r_2}, \quad \beta_2 = -\frac{1}{r_1 r_2}.$$

The reciprocals of the roots and the corresponding parameters β_1 and β_2 are in Table 5.1.

Figure 5.3 shows the coverage probabilities for each process with the nominal coverage probability of 0.9 and $n = 30$. The standard error of the coverage probability was less than 5×10^{-4} in all cases. The Bayesian methods perform much better in all cases. Figures 5.4 and 5.5 show the spectral densities and the associated autocorrelation functions of the corresponding

processes. The spectral densities are scaled such that the total density integrates to 1.

Table 5.1: The AR(2) models used in the simulation experiments.

r_1^{-1}	r_2^{-1}	β_1	β_2
0.9	0.5	1.4	-0.45
0.9	-0.5	0.4	0.45
-0.9	0.5	-0.4	0.45
-0.9	-0.5	-1.4	-0.45
0.5	0.5	1.0	-0.25
0.5	-0.5	0	0.25
$0.9 \exp(\frac{i}{5})$	$0.9 \exp(-\frac{i}{5})$	1.76	-0.81
$0.9 \exp(\frac{i\pi}{2})$	$0.9 \exp(-\frac{i\pi}{2})$	0	-0.81
$0.9 \exp(\frac{i3\pi}{4})$	$0.9 \exp(-\frac{i3\pi}{4})$	-1.27	-0.81

As in the case of AR(1) process, the standard plug-in method gives much smaller coverage probabilities than the Bayesian method. We further see that when the mass of the spectral density function in Figure 5.4 is mostly on the smaller frequencies, corresponding to the slowly decaying autocorrelation function in Figure 5.5, the coverage probabilities stay under the nominal coverage probability. Furthermore, Jeffreys's prior performs slightly better than the other priors. The plug-in method is very poor, especially for longer forecast horizons.

When the mass of the spectral density function is mostly peaked on the medium or high frequencies and the corresponding autocorrelation functions alternate, the forecast horizon does not seem to affect the coverage probabilities neither for the Bayesian nor for the plug-in methods. The Bayesian methods are almost exactly equal to the nominal coverage or somewhat exceed it, while the plug-in method is staying clearly below the nominal coverage. In the case where the spectral density function is very flat, the Bayesian methods give coverage probabilities slightly over the nominal level, while the plug-in method stays below. Assuming stationarity with uniform prior seems to provide coverage probabilities little below those of the uniform and Jeffreys's priors.

6 Annual gross domestic product growth

As an empirical example we applied our method to forecasting the annual gross domestic product (GDP) changes (in percentages) in the United Kingdom (UK) and Spain (Figure 6.1). The data is from World Bank's databank (<http://databank.worldbank.org>). Observations from 1962–2001 are used for finding an appropriate model for each series. Then the series are forecast for 10 years ahead 2002–2011. In the following we first pretend, as far as possible, that we do not know what has happened in the forecast period. Afterwards when future has been revealed, we try to learn from it.

Based on the autocorrelation and partial autocorrelation functions, we assume an AR(1) model for both series. The least squares estimates of the autoregressive coefficients are 0.35 for the UK and 0.65 for Spain. The estimates for β_0 are 1.77 (UK) and 1.25 (Spain) and for σ 1.93 (UK) and 1.83 (Spain).

The residuals of the Spanish series are well behaved. There is neither apparent autocorrelation nor deviation from normality. As for the residuals

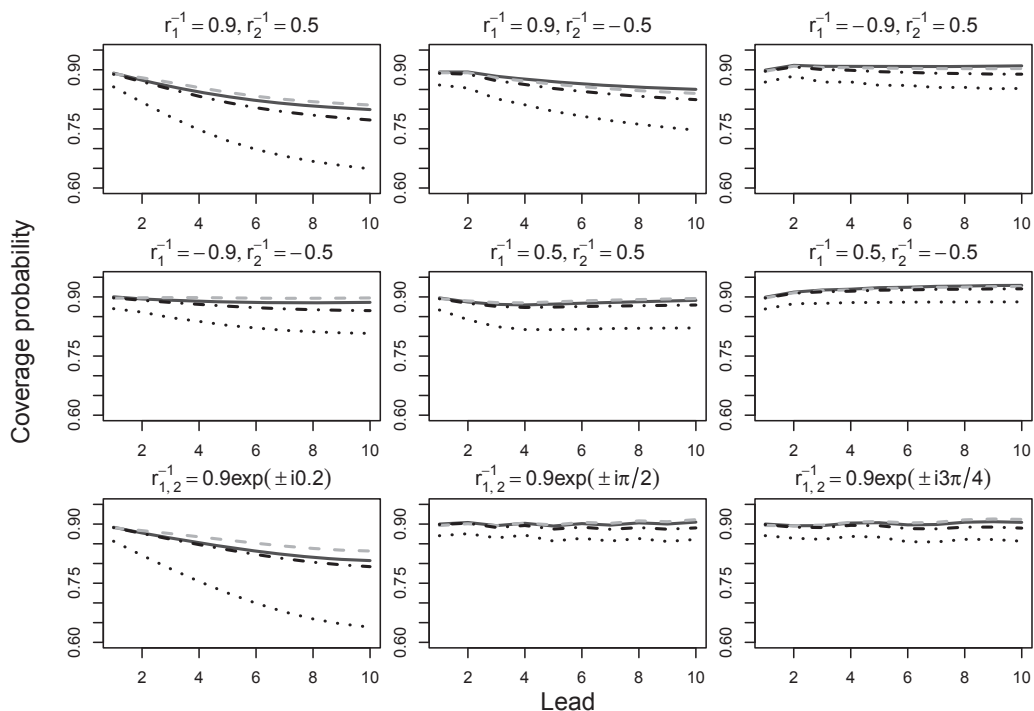


Figure 5.3: The coverage probabilities of the nine AR(2)-processes given in Table 5.1, with $n = 30$. The nominal coverage is 0.9. The black dotted line shows the coverage probabilities of the traditional plug-in method. The other lines are related to the Bayesian methods; the solid dark grey line corresponds to the uniform prior, the light grey dash to Jeffreys's prior, and the dot-and-dash line to the uniform stationary prior.

from the UK series, there appears to be two fairly large negative residuals which stem from the large drops in 1974 and 1980 (i.e. 2 in 40 years). Therefore we should not be surprised to see at least one aberrant value in the forecast period of ten years. No autocorrelation is left in the residuals of the UK series. Granted, we should take into account possible outlying values, but it would require models outside autoregressive family, which is beyond the scope of this article. Therefore we forecast with the estimated model also in this case. In summary, we expect that the Spanish GDP is likely to stay within 90% interval during the forecast period with possibly at most one minor violation. But we are more uncertain whether this holds for the GDP of the UK.

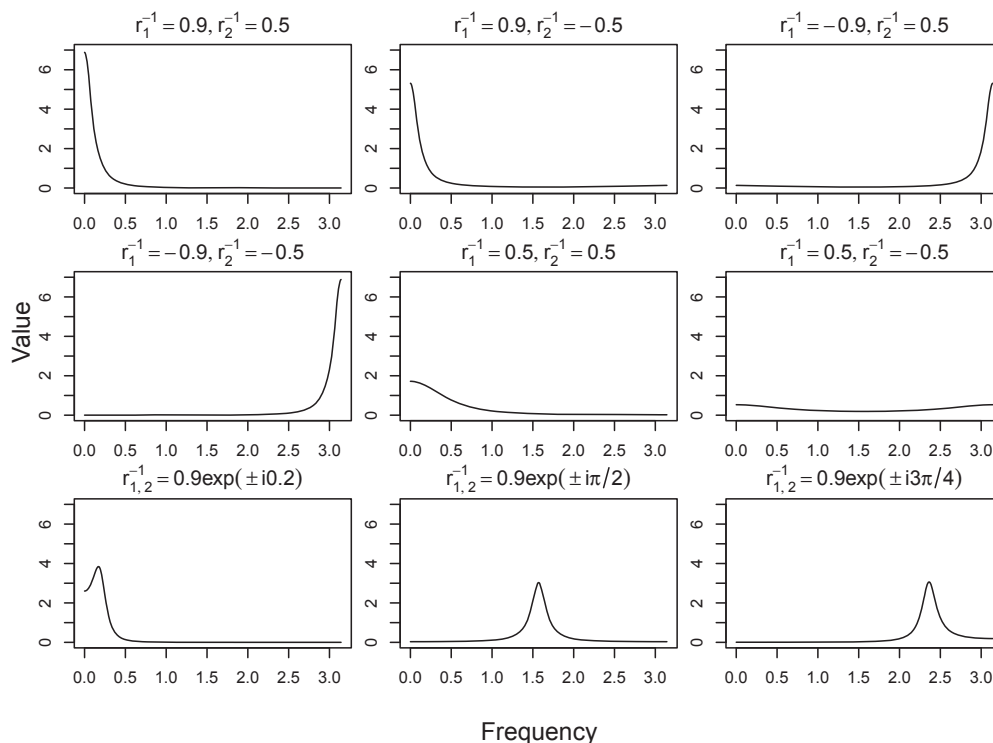


Figure 5.4: The scaled spectral density functions of the nine AR(2)-processes from Table 5.1.

Now, of course, we know that the preceding reasoning is too optimistic. In 2001 there were not any indication of the financial crisis starting in 2008. Figure 6.1 shows the actual values, the 90% prediction intervals and the point predictions for both series. The latter values in the Bayesian forecasting are posterior medians, computed by setting $\alpha = 0.5$. As expected, the Bayesian method produces wider prediction intervals than the standard plug-in method for both series and substantially wider for the Spanish series. We expected at most one minor violation in the Spanish series, whereas we see one large drop below the 90% prediction limit. In the UK series there are one value at the boundary (in 2008) and one significantly below the boundary (in 2009). The former is plausible bearing in mind the drops in 1974 and 1980, whereas latter is exceptionally low. But in both cases the main body of values in 2002–2011 especially those in 2010 and 2011 are well within boundaries. In summary, the values in 2009 are exceptionally low in both

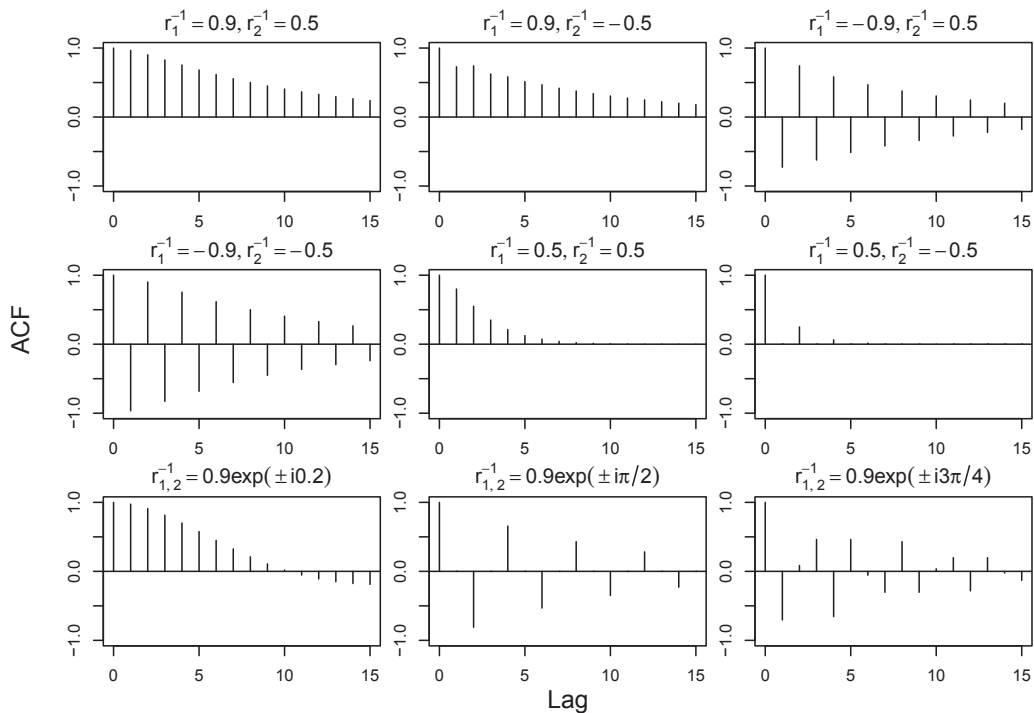


Figure 5.5: The autocorrelation functions of the nine AR(2)-processes from Table 5.1.

series compared to what our model predicts.

There is also one more thing we can learn. Although the differences between the Bayesian intervals are mostly negligible, we see that Jeffreys's prior leads to significantly higher upper limit for the Spanish series. The same is true also for the the point forecasts (i.e. medians). A closer examination of the importance weights using Jeffreys's prior shows that large values are associated with a large values of the autoregressive coefficient. This seems to be due to the large difference between the first observation and the stationary mean, leading to doubts on the stationarity of the series. Our conclusion is that the uniform prior combined with the conditional likelihood is more preferable to Jeffreys's prior in this case.

What should we think of the drops in 2009 occurring in both series? Looking at the series as such and ignoring other information they seem to be single unlikely events rather than the sign of a structural break. There

are at least two arguments for this. Firstly, the values after 2009 (in 2010 and 2011) fit rather well into the main body of the data. Secondly, we computed prediction bands such that the observed values of the year 2009 are lying exactly on the boundary. The corresponding coverage probabilities are $1 - \alpha$ with $\alpha = 0.0016$ for the UK and with $\alpha = 0.0078$ for the Spanish series. Moreover the values in 2009 are also extremely aberrant compared with those in the 40 year period 1962–2011. In summary, the values in 2009 are highly improbable in the light of the rest of the data.

Nevertheless, the accumulated information on world economy and the crisis concerning the euro countries till the end of 2012 makes us think that the assumption of a structural break should still be considered seriously. If we are to forecast from 2012 onwards we should carefully explore the methods suggested by Clements and Hendry (1999).

Table 6.1 gives numeric information on the predictions and prediction intervals related to our application under the assumption that our model is correct. Firstly, it contains the actual prediction limits with their Monte Carlo standard errors obtained from the formulas in section 3.3. Secondly, it shows the coverage probabilities for both models in case they were true. The forecasting horizon is $k = 1, 10$ and the nominal coverage is 0.90. The two Bayesian methods give one-step ahead prediction intervals which are practically correct. When $k = 10$, the coverage probabilities for the UK slightly differ from the nominal ones. The standard method gives intervals which have coverage probabilities below the nominal level in both cases. These comparisons concern the chosen models only, not the actual series: The reported coverage probabilities are averages corresponding to AR(1) models with coefficients 0.35 and 0.65. This explains why the actual intervals for the Spanish series with uniform and Jeffrey’s priors are of considerably different widths, though their coverage probabilities are close to each other on the average. The deviance may also be related to the aberrant starting value.

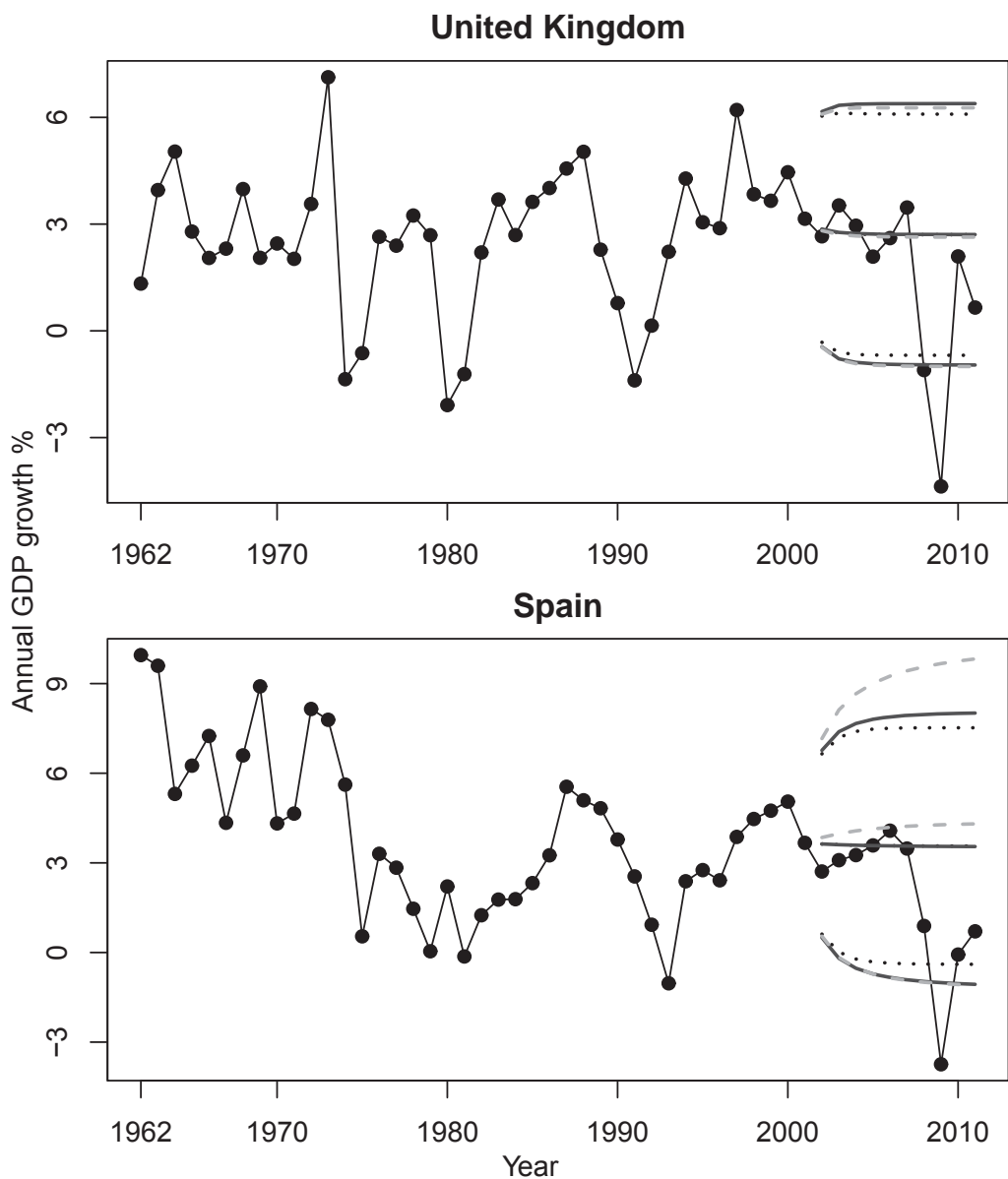


Figure 6.1: The annual GDP 1962–2001 in the UK and Spain together with the point predictions, actual values and 90% prediction intervals for the years 2002–2011. Solid grey lines represents intervals and point estimates computed by the Bayesian method with uniform prior, the grey dashed lines corresponds to the Bayesian method with Jeffreys’s prior and black dotted line corresponds to the standard plug-in method.

Table 6.1: Coverage probabilities and prediction limits, with standard errors, related to the fitted models for the UK and SPA GDP series. The forecast horizon is 2002–2011, and the nominal coverage probability is 0.9.

	Uniform prior		Jeffreys's prior		Plug-in	
	$k = 1$	$k = 10$	$k = 1$	$k = 10$	$k = 1$	$k = 10$
United Kingdom						
Coverage	0.900	0.906	0.900	0.907	0.883	0.881
$\hat{b}_{1-\alpha}$	6.164	6.388	6.089	6.269	6.037	6.091
s.e.($\hat{b}_{1-\alpha}$)	0.002	0.003	0.002	0.003	–	–
\hat{b}_α	-0.448	-0.960	-0.462	-1.003	-0.321	-0.687
s.e.(\hat{b}_α)	0.002	0.003	0.002	0.003	–	–
Spain						
Coverage	0.899	0.892	0.899	0.895	0.881	0.850
$\hat{b}_{1-\alpha}$	6.766	8.014	7.159	9.825	6.647	7.521
s.e.($\hat{b}_{1-\alpha}$)	0.002	0.005	0.003	0.021	–	–
\hat{b}_α	0.498	-1.066	0.546	-1.097	0.616	-0.395
s.e.(\hat{b}_α)	0.002	0.006	0.003	0.011	–	–

We have used the sample size of $N = 100,000$ in our Monte Carlo simulation, and therefore the standard errors are fairly small. Standard errors should decrease at the rate $1/\sqrt{N}$ that is also confirmed by our experiments. With $N = 1000$ the standard errors are approximately tenfold compared to those in Table 6.1.

We have also compared the probabilities that the future value lies below the lower limit or above the upper limit. They seem to be approximately equal. Thus, our method seems to produce equal tail prediction intervals.

7 Discussion

We have shown the benefits of the Bayesian approach to prediction interval calculations under autoregressive schemes. Our message to the practitioners is that there are appropriate prior distributions leading to improved prediction intervals compared to those obtained by the common plug-in method. It has turned out that the uniform and Jeffreys's priors meet most practical goals for such intervals. Jeffreys's prior might have a slight advantage as regards coverage probabilities, although the dependence on the initial observations may have detrimental effects if the starting values are too far from their mean. Our simulation method is straightforward and easy to understand and to implement. An estimate for the Monte Carlo error can also be obtained.

It is plain that when the length of the time series increases, the parameter uncertainty decreases and thus also the coverage probabilities get closer to the nominal level. For example in AR(1) case, with $n = 100$, $h = 10$, and $\beta_1 = 0.5$, the coverage probability of the plug-in method roughly achieves the nominal probability. But any guidelines when to use the simulation method proposed here, instead of the plug-in method, seems to require massive Monte Carlo experimenting. The reason is that the outcome depends on many variables: the length of the series and forecasting horizon, the order of the model and the chosen coefficients. We believe that the proposed approach could be taken as a default method, because it is computationally so light and hardly ever gives results worse than the plug-in method.

The prediction intervals are computed under the Gaussian assumption. We have made some simulation experiments under AR(1) and AR(2) models, where the true errors are from Student's t distribution with 5 degrees of freedom and from Laplace distribution, but we still compute the prediction

intervals under Gaussian assumption. In both cases the coverage probabilities are smaller than in the case where the true errors are Gaussian, but usually the difference is about one percentage point or less. Although such a small experiment does not afford any general conclusions, it makes us to conjecture that the method is fairly robust against minor deviation from normality.

Although we have handled univariate AR processes only, the method could be extended to general ARIMA and vector autoregressive models, but more careful considerations of prior distributions are then needed. In addition, these complex models can be more sensitive to structural breaks and other issues discussed by Clements and Hendry (1999).

Finally, when dealing with a particular data set, some evidence of the adopted prior is obtained by simulating the estimated model as we have done in section 6.

Acknowledgements

The authors thank the editors of the volume and two anonymous referees for their comments and constructive criticism which has led to considerable improvements. We also thank the participants of the Conference in honour of Andrew Harvey's 65 year for the discussion and comments. Jouni Helske thanks for the financial support from Emil Aaltonen Foundation.

References

- C. F. Ansley and R. Kohn. Prediction mean squared error for state space models with estimated parameters. *Biometrika*, 73:467–473, 1986.
- M. Arellano and S. Bonhomme. Robust priors in nonlinear panel data models. *Econometrica*, 77:489–536, 2009.
- O. E. Barndorff-Nielsen and D. R. Cox. Prediction and asymptotics. *Bernoulli*, 2:319–340, 1996.
- R. Beran. Calibrating prediction regions. *Journal of the American Statistical Association*, 85:715–723, 1990.
- J. Berger. The case of objective Bayesian analysis. *Bayesian Analysis*, 1: 385–402, 2006.

- J. O. Berger and R. Yang. Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, 10:461–482, 1994.
- G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, Fourth edition, 2008.
- L. Broemeling and M. Land. On forecasting with univariate autoregressive processes: A Bayesian approach. *Communications in Statistics – Theory and Methods*, 13:1305–1320, 1984.
- C. Chatfield. Calculating interval forecasts. *Journal of Business & Economic Statistics*, 11:121–135, 1993.
- C. Chatfield. Model uncertainty and forecast accuracy. *Journal of Forecasting*, 15:495–508, 1996.
- G.C. Chow. Multiperiod predictions from stochastic difference equations by Bayesian methods. In S. E. Fienberg and A. Zellner, editors, *Studies in Bayesian Econometrics and Statistics*, chapter 8, pages 313–324. North-Holland, Amsterdam, 1974.
- M. P. Clements and D. F. Hendry. *Forecasting Non-stationary Economic Time Series*. The MIT Press, Cambridge, 1999.
- M. P. Clements and J. H. Kim. Bootstrap prediction intervals for autoregressive time series. *Computational Statistics & Data Analysis*, 51:3580–3594, 2007.
- G. Datta and R. Mukerjee. Probability matching priors for predicting a dependent variable with application to regression models. *Annals of the Institute of Statistical Mathematics*, 55:1–6, 2003.
- G. Datta and R. Mukerjee. *Probability Matching Priors: Higher Order Asymptotics*. Springer, New York, 2004.
- D. Firth. Bias reduction of maximum likelihood estimates. *Biometrika*, 80: 27–38, 1993.
- D. A. S. Fraser, N. Reid, E. Marras, and G. Y. Yi. Default priors for Bayesian and frequentist inference. *Journal of the Royal Statistical Society, Series B*, 72:631–654, 2010.

- W. A. Fuller and D. P. Hasza. Properties of predictors for autoregressive time series. *Journal of the American Statistical Association*, 76:155–161, 1981.
- J. Geweke. *Contemporary Bayesian Econometrics and Statistics*. Wiley, Hoboken, 2005.
- J. Geweke and C. Whiteman. Bayesian forecasting. In C. W. J. Elliott, G. Granger and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, chapter 1, pages 3–80. Elsevier B.V., Amsterdam, 2006.
- M. Grigoletto. Bootstrap prediction intervals for autoregressions: Some alternatives. *International Journal of Forecasting*, 14:447–456, 1998.
- A. C. Harvey. *Time Series Models*. Harvester Wheatsheaf, New York, Second edition, 1993.
- P. Kabaila and K. Syuhada. Improved prediction limits for AR(p) and ARCH(p) processes. *Journal of Time Series Analysis*, 29:213–223, 2008.
- J. H. Kim. Bootstrap prediction intervals for autoregression using asymptotically mean-unbiased estimators. *International Journal of Forecasting*, 20: 85–97, 2004.
- S. I. Liu. Multiperiod Bayesian forecasts for AR models. *Annals of the Institute of Statistical Mathematics*, 46:429–452, 1994.
- G. Masarotto. Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, 6:229–239, 1990.
- L. Pascual, J. Romo, and E. Ruiz. Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, 25:449–465, 2004.
- D. Pfeiffermann and R. Tiller. Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26:893–916, 2005.
- P. C. B. Phillips. The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics*, 9:241–261, 1979.
- P. C. B. Phillips. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, 6:333–364, 1991.

- R. Prado and M. West. *Time Series—Modeling, Computation, and Inference*. Chapman & Hall/CRC, Boca Raton, 2010.
- B. Quenneville and A. C. Singh. Bayesian prediction mean squared error for state space models with estimated parameters. *Journal of Time Series Analysis*, 21:219–236, 2000.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- A. Rodriguez and E. Ruiz. Bootstrap prediction intervals in state-space models. *Journal of Time Series Analysis*, 30:167–178, 2009.
- C. A Sims. Comment by Christopher A. Sims on 'To criticize the critics', by Peter C. B. Phillips. *Journal of Applied Econometrics*, 6:423–434, 1991.
- C. A Sims and H. Uhlig. Understanding unit rooters: A helicopter tour. *Econometrica*, 59:1591–1599, 1991.
- R. D. Snyder, J. K. Ord, and A. B. Koehler. Prediction intervals for ARIMA models. *Journal of Business & Economic Statistics*, 19:217–225, 2001.
- P. A. Thompson and R. B. Miller. Sampling the future: A Bayesian approach to forecasting to univariate time series models. *Journal of Business & Economic Statistics*, 4:427–436, 1986.
- P. Vidoni. Improved prediction intervals for stochastic process models. *Journal of Time Series Analysis*, 25:137–154, 2004.
- P. Vidoni. A simple procedure for computing improved prediction intervals for autoregressive models. *Journal of Time Series Analysis*, 30:577–590, 2009.
- A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York, 1971.