Prediction Inference for Time Series

XAVIER DE LUNA

Department of Economics, Umeå University,
90187 Umeå, Sweden

ABSTRACT

In this paper we briefly review Bayesian and frequentist prediction inference for time series, and then advocate the use of guaranteed-content prediction intervals. These intervals are such that their content (or coverage) is guaranteed with a given high probability. They, thus, are more relevant for the observed time series at hand than classical prediction intervals, whose content is guaranteed merely on average over hypothetical repetitions of the prediction process. Guaranteed-content intervals should, therefore, conciliate Bayesians and frequentists when no prior belief on the parameterization is available. This type of prediction inference has, however, been ignored in the time series context because of a lack of results. This gap is filled by deriving asymptotic results for a general family of autoregressive models, thereby extending existing results in non-linear regression. The actual construction of guaranteed-content prediction intervals directly follows from this theory. Simulated and real data are used to illustrate the practical difference between classical and guaranteed-content prediction intervals for ARCH models.

KEYWORDS Guaranteed content interval; heteroskedasticity; non-linear autoregression; parameter uncertainty

INTRODUCTION

The purpose of this paper is to advocate an approach to prediction inference for time series which is based on prediction intervals with guaranteed content. By prediction inference we mean the analysis of a set of data with the aim of obtaining the forecast of a future event. Ideally, prediction intervals are subsets of the space of all possible events which contain the future and hence unobserved event with a given probability, usually called content or coverage. Such a definite probabilistic statement is unfortunately rarely available in practice unless the Bayesian approach is embraced. However, it is sometimes possible to obtain a prediction interval whose content is guaranteed with a given high probability. It is shown in this paper how to construct such guaranteed-content prediction intervals for a large family of parametric models for time series; examples include threshold autoregression (TAR), feed-forward neural network, autoregressive conditional heteroskedasticity (ARCH) and many others, described for instance in Granger and Teräsvirta (1993).

The guarantee of the content with a probabilistic statement is to be put in contrast with classical (frequentist) prediction intervals whose content is guaranteed only in average over hypothetical repetitions of the prediction process. In practice, however, merely a single time series is usually available. The average meaning of the classical prediction interval is moreover not always acknowledged by the non-expert who often has a blind confidence in the content. This pitfall can be avoided with the Bayesian approach to prediction inference, which allows for the computation of the probabilistic distribution of the future event of interest. Bayesian prediction inference is, however, subject to some practical difficulties. For instance, to date no general-purpose procedure exists to compute the predictive distribution (the distribution of the future event) because case dependent Monte Carlo techniques are often needed. A growing literature results from this fact, see e.g. Geweke (1989b), Geweke and Terui (1991), and Polasek and Kozumi (1996).

Guaranteed-content prediction intervals are not new and an history can be found in Aitchison and Dunsmore (1975, p. 128). Wald and Wolfowitz (1946) were the first to consider them for the Normal distribution and independent observations. More recently Carroll and Ruppert (1991) gave asymptotic results in the non-linear regression setting, and we show in this article under which conditions Carroll and Ruppert's results may be extended to non-linear autoregression processes.

The article is organized as follows. Section 2 introduces the models for time series considered in the study. Sections 3 and 4 give an overview of, respectively, Bayesian and frequentist prediction inference within the time series context of interest; the strengths and weaknesses of both approaches are pointed out. Section 5 introduces the guaranteed-content prediction intervals and gives asymptotic results of practical use. ARCH simulated and real data illustrate the use of the prediction intervals. A discussion concludes the paper in Section 5. Proofs are gathered in the Appendix.

MODELS FOR TIME SERIES

Assume a real valued time series, $x_1, x_2, \ldots, x_n$, has been observed. A common (some would say the main) inferential problem is to provide a forecast of a future unobserved value, say x_{n+h}. Thus, the probabilistic object of interest is (if it exists) the conditional density of x_{n+h} given $x_1, \ldots, x_n$, denoted $p(x_{n+h}|x_n, \ldots, x_1)$. This density is generally unknown and need to be deduced from the data. To narrow down the problem it is customary to use a parametric specification general enough to provide a good approximative working tool. In our case we consider a parametric model which is widely used because of its direct interpretation in terms of first and second order conditional moments. We thus assume that, for $p \geqslant 0$ fixed,

$$x_{t+h}|y_{t,t}, \xi \sim \mathcal{L}\big\{f(y_{t,p};\theta), g(y_{t,p};\eta)^2\big\}, \tag{1}$$

where the notation $y_{t,p} = (x_t, \ldots, x_{t-p+1})'$ and $\xi' = (\theta', \eta')$ is used. We also use $\mathcal{L}$ to denote any given distribution law, here with expectation $f(y_{t,p};\theta)$ and standard deviation $g(y_{t,p};\eta)$; both are real valued functions depending implicitly on

h. The prediction intervals discussed in Section 4 and 5 are mainly directed at unimodal laws. In practice, the symmetric Normal distribution is often use for $\mathcal{L}$. The above model is mainly appropriate for equally spaced observations. It is intuitively appealing because, for a given parameter value ξ, the conditional expectation $f(y_{t,p};\theta)$ is the optimal (in a mean squared error sense) point forecast for x_{n+h} and $g(y_{n,p};\eta)$ measures its variability which is allowed to vary with time. Note that at this stage no stationary assumption is made, although classical stationary autoregressive processes are important special cases of (1), examples of which are given in the Introduction. Polynomial functions are an example of family of simple non-linear functions useful to model $f(y_{t,p};\theta)$ and $g(y_{t,p};\eta)^2$ without relying on a stationary model for the generating stochastic process, see de Luna (1998).

The parameter ξ is most of the time a modeling artifact which eases the primary task of computing or approximating the density $p(x_{n+h}|x_n,\ldots,x_1)$. There are two main schools when it comes to deal with a parameter. It can be considered as a random variable (Bayesian school) or as having a unique but unknown fixed value (frequentist school). In the two next sections we review these two approaches to prediction inference in time series and point out their respective strengths and weaknesses. Note that, in this article, we try to avoid to enter into the philosophical controversies existing between Bayesians and frequentists, concerning, for instance, the definition of probabilities. We adopt instead a pragmatic view in the exposition of these two different approaches to inference.

BAYESIAN PREDICTION INFERENCE

The Bayesian prediction inference is conceptually simple. Because the parameter ξ is here a random variable, we can associate to it a prior distribution, with say density $p(\xi)$. This density must summarize the beliefs on the parameter, independently of the data collected. These beliefs are then updated with the observation of $x_1,\ldots,x_n$, thereby giving the so called posterior density

$$p(\xi|y_{n,n}) \propto p(y_{n,n}|\xi)p(\xi). \tag{2}$$

Note that $p(y_{n,n}|\xi)$ is the likelihood function for the parameter ξ. The prediction density can then be obtained as follows

$$\begin{aligned} p(x_{n+h}|y_{n,n}) &= \int p(x_{n+h},\xi|y_{n,n})d\xi \\ &= \int p(x_{n+h}|y_{n,n},\xi)p(\xi|y_{n,n})d\xi. \end{aligned} \tag{3}$$

The prediction distribution is therefore an average of the parametric model predictive distribution (1) over the posterior distribution of the parameter. This so far elegant approach suffers from two main drawbacks when it has to be actually implemented: (i) A prior distribution is needed for a parameter that has often little intuitive interpretation. If no such prior is available, then some ad hoc solutions are often used such as taking the posterior proportional to the likelihood, i.e. using

the improper prior $p(\xi) = 1$. The prediction density is in this case not justified as such even if the posterior is itself a proper distribution. (ii) The integral (3) seldom has a known closed form in which case simulation techniques need to be used. For instance, assuming that you are able to draw ξ values from the posterior distribution, (3) can be estimated either by averaging $p(x_{n+h}|y_{n,n},\xi)$ over the simulated values for ξ, or by further drawing from $p(x_{n+h}|y_{n,n},\xi)$ for each ξ value obtained from the posterior distribution. In the latter, the obtained simulated values allow for a classical empirical estimation of $p(x_{n+h}|y_{n,n})$. However, drawing from the posterior distribution may not be straightforward, for instance, when the constant of proportionality in (2) is unknown. Complex numerical or stochastic methods such as Markov chain Monte Carlo need then to be used, see e.g. Geweke (1989a) and Gelman *et al.* (1995).

Some will argue that problems encountered with (i) are often due to an inadequate model. A typical Bayesian model should be parameterized in an interpretable manner so as to enable the scientist to form his/her prior beliefs on the parameter. In time series the Dynamical Linear Models of West and Harrison (1989) are of this type. These models are, however, intrinsically linear. The second point (ii) is mainly a practical issue in which, nowadays, very much research effort is carried over. However, no general methodology is available at present and different models for stochastic processes have to be studied in a case by case basis. Recent works on special cases of (1) include Geweke (1989b) on ARCH models and Geweke and Terui (1991) on TAR. Note that a non-parametric approach allows for a more general approach (see Müller *et al.*, 1997), but this is out of the scope of this paper.

FREQUENTIST PREDICTION INFERENCE

The frequentist approach to prediction inference, although often simple to implement, is conceptually intricate. This is because the prediction density −here $p(x_{n+h}|y_{n,n},\xi)$− is generally not computable, the value of ξ being unknown. To circumvent this problem the notion of *prediction interval* (sometimes called *tolerance interval*) must be introduced. These intervals cannot be based on the unknown prediction density and, therefore, a frequentist argument need to be used.

Prediction intervals are constructed with the help of a pivotal quantity which in our case may be chosen as

$$S(\xi) = \frac{x_{n+h} - f(y_{n,p};\theta)}{g(y_{n,p};\eta)}.$$

To simplify the exposition we assume in this section that $\mathcal{L}$ is the Normal distribution. Thus, $S(\xi)$ has N(0,1) distribution when conditional on $y_{n,p}$, for a given ξ. Hence, the interval

$$\mathbb{P}_\alpha(\xi) = [f(y_{n,p};\theta) \pm g(y_{n,p};\eta)\Phi^{-1}_{1-\alpha/2}],$$

where $\Phi^{-1}_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard Normal distribution, has the property

$$\Pr_\xi\{x_{n+h} \in \mathbb{P}_\alpha(\xi)|y_{n,p}\} = 1-\alpha.$$

The probability $1-\alpha$ is called the *content* or *coverage* of the interval.

Suppose now that an estimator $\hat{\xi} = \xi(y_{n,n})$ for ξ is available. It is then usual practice to use the plug-in prediction interval $\mathbb{P}_\alpha(\hat{\xi})$. Under regularity conditions we have

$$\mathrm{E}_\xi\left[\Pr_\xi\{x_{n+h} \in \mathbb{P}_\alpha(\hat{\xi})|y_{n,p},\hat{\xi}\}\big|y_{n,p}\right] = 1-\alpha+O_p(n^{-1}). \tag{4}$$

The two operators E and Pr are used instead of a single one for better readability. We use throughout the Mann and Wald (1943) notation, o_p and O_p, for orders of convergence in probability. The above result, which is justified in the Appendix, leads us to call $\mathbb{P}_\alpha(\hat{\xi})$ a mean-content prediction interval. This type of intervals is of common use in practice and we also refer to it as classical prediction interval. Prediction inference based on $\mathbb{P}_\alpha(\hat{\xi})$ neglects the variability due to the estimation of ξ. We see from (4) that this uncertainty influences the mean-content with an order $O_p(n^{-1})$. The use of techniques to improve on the convergence rate of the mean-content is called *calibration*. Analytical calibration can be undertaken, for instance, by calculating higher order terms in the asymptotic approximation (4) and modifying α in consequence (Cox, 1975, Barndorff-Nielsen and Cox, 1996). However, this is feasible only in simple cases. Bootstrap is an usual remedy to the analytical intractability because it allows for high order approximation of expectations by using replicates from the actual time series, see e.g. Beran (1990). Note, however, that the expectation of interest in (4) is conditional on the last p observed values, $y_{n,p}$, of the times series. Stine (1987) and Kabaila (1993) have proposed a bootstrap scheme to construct prediction intervals for linear autoregressions. Their method is based on bootstrap copies of the original time series which are generated backward in time and conditioned on $y_{n,p}$. At this moment, we are not aware of how this could be generalized to the non-linear processes considered in this paper. Other simulation techniques have been proposed in the literature in order to take into account parameter uncertainty, see Breidt et al. (1995). Finally, prediction intervals for misspecified linear models were studied in de Luna (2000). Here again, the analytical developments are difficult to generalize to non-linear situations.

GUARANTEED-CONTENT PREDICTION INTERVALS

In the two previous sections, we have pinpointed possible drawbacks of Bayesian and classical frequentist prediction inference when applied to non-linear time series models described by (1). We now propose another type of prediction inference for time series. When no prior beliefs on the parameter is available and if the prediction inference is carried out on a single realization of a stochastic process, it is more appropriate to construct prediction intervals whose content, $1-\alpha$, is

guaranteed with high probability rather than merely in average. Formally, we want $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi}) \subset \mathbb{R}$ such that

$$\Pr_{\xi}\big[\Pr_{\xi}\{x_{n+h} \in \mathbb{G}_{\alpha,\gamma}(\widehat{\xi})|\widehat{\xi}, y_{n,p}\} \geq 1-\alpha\big|y_{n,p}\big] = 1-\gamma. \tag{5}$$

We say in the sequel that $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi})$ is a guaranteed-content prediction interval. Before to explicitly construct $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi})$, we state some regularity conditions. In the sequel, we denote Ψ the standardized, i.e. with mean zero and variance one, distribution function associated to $\mathcal{L}$.

(H1) The distribution Ψ possesses a well behaved density ψ.

(H2) $f(\cdot;\theta)$ and $g(\cdot;\eta)$ are twice continuously differentiable with respect to θ and η respectively; with second derivative bounded in probability.

(H3) $E(\widehat{\xi}-\xi_0) = O(n^{-1})$, $E\{(\widehat{\xi}-\xi_0)(\widehat{\xi}-\xi_0)'\} = O(n^{-1})$, where $\widehat{\xi}$ is an estimator of ξ_0, the single limit value of interest. In particular, model (1) needs to be valid only for $\xi = \xi_0$.

(H4) $\sqrt{n}(\widehat{\xi}-\xi_0) \overset{n\to\infty}{\sim} N(0,\Sigma)$ (asymptotic normality).

(H5) $\sqrt{n}\widehat{\xi}$ and $y_{n,p}$ are asymptotically independent.

Assumptions (H3-4) are commonly met by maximum likelihood and alike estimation procedures. The last assumption (H5) is not always trivial to check. However, a sufficient condition is that the stochastic process has some vanishing dependence structure (e.g., mixing conditions, see Doukhan, 1994). Finally, the smoothness assumptions of (H2) can be weakened from case to case.

A guaranteed-content prediction interval $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi})$ can be constructed by tuning α in the plug-in interval, introduced in the previous section for the Normal model. The naive choice $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi}) = \mathbb{P}_{\alpha}(\widehat{\xi})$ is not satisfactory as we see with the results given below. We start by addressing the case of one-sided intervals.

PROPOSITION. *Let (H1-5) hold. Then, for*

$$\mathbb{G}_{\alpha,\gamma}(\widehat{\xi}) = (-\infty, f(y_{n,p};\widehat{\theta}) + g(y_{n,p};\widehat{\eta})\Psi^{-1}_{1-\beta}],$$

where $\Psi^{-1}_{1-\beta}$ denotes the $(1-\beta)$-quantile of the distribution function Ψ, (5) holds up to an order $o_p(n^{-1/2})$ if

$$\beta = \alpha + n^{-1/2}\Phi^{-1}_{\gamma}\psi(\Psi^{-1}_{1-\beta})(d'\Sigma d)^{1/2}, \tag{6}$$

where $d' = (g(y_{n,p};\eta_0)^{-1}f_0', g(y_{n,p};\eta_0)^{-1}\Psi^{-1}_{1-\beta}g_0')$ with $f_0 = \partial f(y_{n,p};\theta_0)/\partial\theta$ and $g_0 = \partial g(y_{n,p};\eta_0)/\partial\eta$.

See the Appendix for a proof. Equation (6) shows that the parameter uncertainty influences the guaranteed content with an order $O_p(n^{-1/2})$. For the mean coverage (4) this uncertainty was only present with order $O_p(n^{-1})$. Two-sided intervals can be straightforwardly obtained from the Proposition. An interesting simplification arises when the distribution $\mathcal{L}$ is symmetric as it is the case with the Normal law.

COROLLARY. *Let (H2-5) hold and $\mathcal{L}$ be the Normal distribution law. Then, for*

$$\mathbb{G}_{\alpha,\gamma}(\widehat{\xi}) = [f(y_{n,p};\widehat{\theta}) \pm g(y_{n,p};\widehat{\eta})\Phi^{-1}_{1-\beta/2}], \tag{7}$$

(5) holds up to an order $o_p(n^{-1/2})$ *if*

$$\beta = \alpha + n^{-1/2} 2\Phi_\gamma^{-1}\phi(\Phi^{-1}_{1-\beta/2})(d_2'\Sigma_2 d_2)^{1/2}, \tag{8}$$

where ϕ *is the Normal density function,* $d_2 = g(y_{n,p};\eta_0)^{-1}\Phi^{-1}_{1-\beta/2} g_0$ *and* $\Sigma_2 \overset{n\to\infty}{=} nVar(\widehat{\eta})$.

See the Appendix for a proof. Thus, in the Normal case, the parameter uncertainty due to the estimation of θ is less relevant than the one due to the scale parameter η. In this particular case, taking into account the location uncertainty would need higher order approximations. On the other hand, the ambition here is not primarily to take into account parameter uncertainty, but to obtain a prediction inference more relevant for the observed realisation of the stochastic process by ensuring a high guarantee $1-\gamma$. The results above show, however, that these two issues are intrinsically related and that increasing the sample size may not help to get ride of parameter uncertainty if $1-\gamma$ is increased at the same time (Φ_γ^{-1} may be arbitrarily large). Finally, note that choosing $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi}) = \mathbb{P}_\alpha(\widehat{\xi})$, i.e. $\beta = \alpha$ in (7), corresponds to $\gamma = 0.5$ if ignoring terms of order $o(n^{-1/2})$ and higher, that is only 50% of the times is the content $1-\alpha$ reached.

We now give some examples to illustrate the Normal case, where the scale parametrization is the most sensitive one. The modelling of the scale variation has for instance been emphasised in financial applications.

Example (homoskedasticity) Setting $g(y_{n,p};\eta) = \eta$, a strictly positive real value, (8) becomes

$$\beta = \alpha + 2\Phi_\gamma^{-1}\phi(\Phi^{-1}_{1-\beta/2})\Phi^{-1}_{1-\beta/2}\eta^{-1}\sqrt{Var(\widehat{\eta})}.$$

Example (ARCH conditional variance) This second example considers Engle's (1982) ARCH modeling of the conditional variance, which corresponds to setting

$$g(y_{t,p};\eta) = (\eta_0 + \sum_{i=1}^{p}\eta_i x^2_{t-i+1})^{1/2},$$

where $\eta = (\eta_0,\ldots,\eta_q)'$. Denoting $v_n = (1, x_n^2,\ldots,x^2_{n-p+1})'$, (8) becomes

$$\beta = \alpha + \Phi_\gamma^{-1}\phi(\Phi^{-1}_{1-\beta/2})\Phi^{-1}_{1-\beta/2}(\eta' v_n)^{-1}(v_n' Var(\widehat{\eta}) v_n)^{1/2}. \tag{9}$$

In the two above examples, results are not dependent on the conditional expectation $f(y_{n,p};\theta)$.

Example (simulated data) In order to illustrate the possible impact of the above results on finite sample situations we use simulated data. We consider the stationary ARCH(6) model, i.e. (1) with $f(y_{t,p},\theta) = 0$, and $g(y_{t,p=6},\eta) = (0.1 + \sum_{i=0}^{4} 0.1 x^2_{t-i} + 0.3x^2_{t-5})^{1/2}$. The Normal distribution is used for $\mathcal{L}$. Three independent time series of length $n = 60, 200, 300$ were simulated (using the random generator of the *Splus* package). Prediction intervals $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi})$ from the Corollary with (9) were constructed for $h = 1$, with $\gamma = 50\%$ (corresponds to classical mean-content interval) and $\gamma =$

TABLE 1. Upper Bounds of Mean-Content ($\gamma = 50\%$) and Guaranteed-Content ($\gamma = 10\%$) prediction Intervals

	n=60		n=100		n=200		n=300	
γ	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1
$\alpha = 0.05$	1.06	1.27	0.83	0.91	2.15	2.43	1.24	1.35
$\alpha = 0.5$	0.36	0.48	0.28	0.32	0.74	0.86	0.43	0.47

Notes: Results are for $h = 1$ and are based on simulated ARCH time series of length $n = 60, 100, 200$ and 300. The intervals whose upper bound is reported are symmetric and centered in zero. The values $\alpha = 5\%$ and 50% correspond to contents of 95% and 50% respectively.

10%. Parameters were estimated by conditional maximum likelihood. Two values of α were considered, namely 5% and 50%. Note that 95% content intervals are usually so wide that alone they are not always interesting for the practitioner (see Granger, 1996). The prediction intervals constructed are symmetric and centered in zero and, therefore, only the computed upper bound of the intervals are displayed in Table 1. The figures in the table show that the relative difference in length between prediction intervals with $\gamma = 50\%$ (mean-content) and $\gamma = 10\%$ may attain 20% for time series of short length $n = 60$. This difference remains important (almost 10%) when $n = 300$, which is a fairly large sample size.

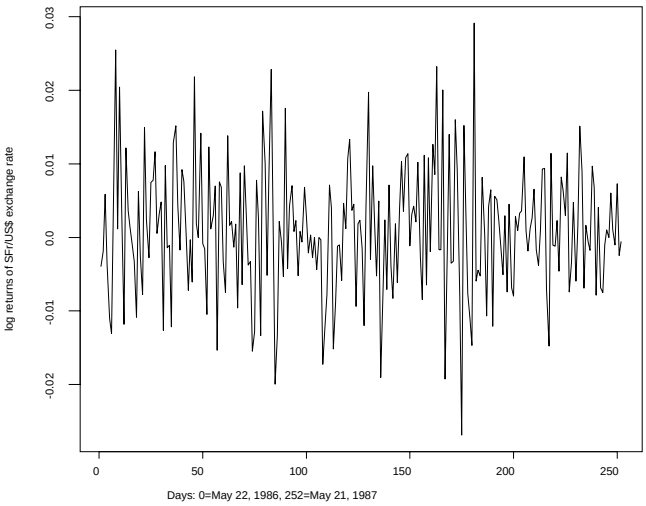

FIGURE 1. Logarithm of relative change in the SFr/US$ exchange rate. Daily data from May 21, 1986 until May 21, 1987.

Example (SFr/US$ exchange rate) We consider here daily data from the exchange rate of the Swiss Franc against the US Dollar. The time series ranges from May 21, 1986 to May 21, 1987. By taking the first differences of the natural logarithm

we obtain a time series of 252 observations. Figure 1 displays a time plot of the transformed data. After exploring the sample (partial) autocorrelation structure of the raw and squared time series, we decided to model the time series as an ARCH(6) process. Parameters were fitted by maximizing the conditional Gaussian maximum likelihood as for the above simulated data. An analysis of the residuals did not contradict the model assumptions in (1), i.e. normality and independence. The constructed prediction intervals for May 22, 1987, are for $\gamma = 50\% : [0 \pm 0.014]$ for $\alpha = 5\%$ and $[0 \pm 0.0048]$ for $\alpha = 50\%$; for $\gamma = 10\%$ we found: $[0 \pm 0.015]$ for $\alpha = 5\%$ and $[0 \pm 0.0052]$ for $\alpha = 50\%$. The relative differences between the two types of prediction intervals are about 7 to 8 percent which obviously matters when the decision-maker has to take a position on the basis of the forecasted volatility; here the range of the interval.

CONCLUSION

Mean-content prediction intervals have often been criticized for being too narrow because they neglect the sources of uncertainty created by the model selection and the parameters estimation stages, see Chatfield (1993). We argue here that it is more judicious to start by acknowledging the inadequacy of their average meaning, and use instead guaranteed-content intervals, for which the parameter uncertainty happen to be more important a factor.

Guaranteed-content intervals were, for example, recently advocated in a non-linear regression setting by Carroll and Ruppert (1991). The use of guaranteed-content prediction interval (or tolerance interval) in place of the classical mean-content version constitutes a real improvement to prediction inference. This is hoped to become more widely acknowledged in time series and more generally in econometrics practice. With this paper, a gap existing in the theory of prediction inference for time series has thus been filled, and, for a general family of parametric models for stochastic processes, guaranteed-content prediction intervals have been constructed. These are wider than classic mean-content prediction intervals because they are more relevant to the time series at hand. We have illustrated this issue with simulated and real data.

Bayesian prediction inference also has the advantage of addressing the actually observed time series by fully taking into account the parameter uncertainty through the posterior density. However, when there is no prior beliefs on the parameterization, we believe guaranteed-content prediction intervals to be more appropriate.

For the Normal model, the results presented become particularly relevant for the scale parameterization. In these cases, conditional heteroskedastic models play more important a role in the distinction between mean- and guaranteed-content prediction intervals. Finally, in practice, when the Normal assumption is not valid, it is common to use a semi-parametric approach (where the distribution $\mathcal{L}$ in (1) is replaced by an empirical distribution). The result of the Proposition is then still

of practical use by replacing the quantile and the density values by their sample counterparts.

ACKNOWLEDGEMENTS

The author is grateful to Kurt Brännäs, Sara Sjöstedt, and Kostas Skouras for their valuable comments on previous versions of this paper. This work was financially supported by the Swedish Institute.

APPENDIX

Proof of the Proposition In the sequel we use Pr instead of $\Pr_{\xi_0}$.With two Taylor expansions, using (H2),

$$\begin{aligned}
&\Pr\{x_{n+h} \leq f(y_{n,p};\widehat{\theta}) + g(y_{n,p};\widehat{\eta})\Psi_{1-\beta}^{-1}|y_{n,p},\widehat{\theta},\widehat{\eta}\} \\
&\quad = \Pr[S(\xi_0) \leq g(y_{n,p};\eta_0)^{-1} f_0'(\widehat{\theta}-\theta_0) \\
&\qquad +\{1+g(y_{n,p};\eta_0)^{-1} g_0'(\widehat{\eta}-\eta_0)\}\Psi_{1-\beta}^{-1} + O_p(n^{-1})|y_{n,p},\widehat{\theta},\widehat{\eta}] \\
&\quad = \Psi[g(y_{n,p};\eta_0)^{-1} f_0'(\widehat{\theta}-\theta_0) \\
&\qquad +\{1+g(y_{n,p};\eta_0)^{-1} g_0'(\widehat{\eta}-\eta_0)\}\Psi_{1-\beta}^{-1} + O_p(n^{-1})].
\end{aligned}$$

Again, with an order one expansion, using (H1),

$$\begin{aligned}
&\Pr\{x_{n+h} \leq f(y_{n,p};\widehat{\theta}) + g(y_{n,p};\widehat{\eta})\Psi_{1-\beta}^{-1}|y_{n,p},\widehat{\theta},\widehat{\eta}\} \\
&\quad = 1-\beta+\psi(\Psi_{1-\beta}^{-1})\{g(y_{n,p};\eta_0)^{-1} f_0'(\widehat{\theta}-\theta_0) \\
&\qquad +\Psi_{1-\beta}^{-1} g(y_{n,p};\eta_0)^{-1} g_0'(\widehat{\eta}-\eta_0)\} + O_p(n^{-1}). \qquad \text{(A.1)}
\end{aligned}$$

Finally, replacing β by $\alpha + n^{-1/2}\Phi_\gamma^{-1}\psi(\Psi_{1-\beta}^{-1})(d'\Sigma d)^{1/2}$ one obtains the result

$$\Pr\left[\Pr\{x_{n+h} \in \mathbb{G}_{\alpha,\gamma}(\widehat{\xi})|\widehat{\xi},y_{n,p}\} \geq 1-\alpha\middle|y_{n,p}\right] = 1-\gamma+o_p(n^{-1/2})$$

for the interval $\mathbb{G}_{\alpha,\gamma}(\widehat{\xi})$ of the Proposition, because of the asymptotic normality of the estimates, (H4), and by assumption (H5).

Proof of Equation (4): This result is a consequence of (A.1) for $\beta = \alpha$, conjugated with (H2), (H3) and (H5).

Proof of the Corollary Consider (A.1), replacing β with $\beta/2$ and Ψ with the standard Normal distribution function Φ. Moreover, the similar result

$$\begin{aligned}
&\Pr\{x_{n+h} \geq f(y_{n,p};\widehat{\theta}) + g(y_{n,p};\widehat{\eta})\Phi_{\beta/2}^{-1}|y_{n,p},\widehat{\theta},\widehat{\eta}\} \\
&\quad = \beta/2+\phi(\Phi_{\beta/2}^{-1})\{g(y_{n,p};\eta_0)^{-1} f_0'(\widehat{\theta}-\theta_0) \\
&\qquad +\Phi_{\beta/2}^{-1} g(y_{n,p};\eta_0)^{-1} g_0'(\widehat{\eta}-\eta_0)\} + O_p(n^{-1})
\end{aligned}$$

applies. Noting that $\phi(\Phi_{\beta/2}^{-1}) = \phi(\Phi_{1-\beta/2}^{-1})$ we have

$$\begin{aligned}
&\Pr\{x_{n+h} \in \mathbb{G}_{\alpha,\gamma}(\widehat{\xi})|y_{n,p},\widehat{\xi}\} \\
&\quad = 1-\beta+2\phi(\Phi_{\beta/2}^{-1})\Phi_{1-\beta/2}^{-1} g(y_{n,p};\eta_0)^{-1} g_0'(\widehat{\eta}-\eta_0) + O_p(n^{-1}).
\end{aligned}$$

As in the Proposition proof, the assumption $\sqrt{n}(\widehat{\eta}-\eta_0) \overset{n\to\infty}{\sim} N(0,\Sigma_2)$, (H4), and hypothesis (H5) complete the proof.

REFERENCES

Aitchison, J. and Dunsmore, I. R., *Statistical Prediction Analysis.* Cambridge: Cambridge University Press, 1975.

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

Beran, R., 'Calibrating Prediction Regions', *Journal of the American Statistical Association,* **85** (1990), 715-723.

Breidt F. J., Davis, R. A. and Dunsmuir, W. T. M., 'Improved Bootstrap Prediction Intervals for Autoregressions', *Journal of Time Series Analysis,* **16** (1995), 177-200.

Carroll, R. J. and Ruppert, D., 'Prediction and Tolerance Intervals with Transformation and/or Weighting', *Technometrics*, **33** (1991), 197-210.

Chatfield, C., 'Calculating Interval Forecasts' (with Discussion), *Journal of Business and Economic Statistics,* **11** (1993), 121-144.

Cox, D. R., 'Prediction Intervals and Empirical Bayes Confidence Intervals', in Gani, J. (ed), *Perspectives in Probability and Statistics*, London: Academic Press, 1975, 47-55.

de Luna, X., 'Projected Polynomial Autoregression for Prediction of Stationary Time Series', *Journal of Applied Statistics*, **25** (1998), 763-776.

de Luna, X., 'Prediction Intervals Based on Autoregression Forecasts', *Journal of the Royal Statistical Society Series D,* (2000), to appear.

Doukhan, P., *Mixing: Properties and Examples.* Berlin: Springer-Verlag, 1994.

Engle, R. F., 'Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation', *Econometrica*, **50** (1982), 987-1007.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B., *Bayesian Data Analysis.* London: Chapman and Hall, 1995.

Geweke, J., 'Bayesian Inference in Econometrics Models Using Monte Carlo Integration', *Econometrica*, **57** (1989a), 1317-1339.

Geweke, J., 'Exact Predictive Densities for Linear Models with ARCH Disturbances', *Journal of Econometrics*, **40** (1989b), 63-86.

Geweke, J. and Terui, N., 'Bayesian Threshold Autoregressive Models for Nonlinear Time Series', *Journal Time Series Analysis*, **14** (1993), 441-454.

Granger, C. W. J., 'Can we Improve the Perceived Quality of Economic Forecasts?', *Journal of Applied Econometrics*, **11** (1996), 1-27.

Granger, C. W. J. and Teräsvirta, T., *Modelling Nonlinear Economic Relationships.* Oxford: Oxford University Press, 1993.

Kabaila, P., 'On Bootstrap Predictive Inference for Autoregressive Processes', *Journal of Time Series Analysis*, **14** (1993), 473-484.

Mann, H. B. and Wald, A., 'On Stochastic Limit and Order Relationships', *Annals of Mathematical Statistics,* **14** (1946), 217-226.

Müller, P., West, M. and MacEachern S., 'Bayesian Models for Non-Linear Autoregressions', *Journal of Time Series Analysis*, **18** (1997), 593-614.

Polasek, W. and Kozumi, H., 'The VAR-VARCH Model: A Bayesian Approach', in Lee, J. C., Johnson, W. O. and Zellner, A. (eds), *Modelling and Prediction: Honoring Seymour Geisser*, Berlin: Springer-Verlag, 1996, 402-422.

Stine, R. A., 'Estimating Properties of Autoregressive Forecasts', *Journal of the American Statistical Association,* **82** (1987), 1072-1078.

Wald, A. and Wolfowitz, J., 'Tolerance Limits for a Normal Distribution', *Annals of Mathematical Statistics*, **17** (1946), 208-215.

West, M. and Harrison, J., *Bayesian Forecasting and Dynamical Models*. Berlin: Springer-Verlag, 1989.