Prediction Intervals for NAR Model Structures Using a Bootstrap Method

De Brabanter, J. , Pelckmans, K., Suykens, J. and Vandewalle J.

KuLeuven

Abstract— We consider the problem of constructing nonparametric prediction intervals for a NAR model structure. Our approach relies on the external bootstrap procedure [**?**]. This method is contrasted with a more traditional approach relying on the Gaussian strategy.

1. Introduction

A great deal of data in business, economics, engineering and the natural sciences occur in the form of time series where observations are dependent. Linear time series models provide powerful tools for analyzing time series data when the models are correctly specified. However, any parametric models are at best only an approximation to the true stochastic dynamics that generates a given data set. Linear time series models are generally the starting point for modeling time series data.

Many data in applications (e.g., sunspot, lynx and blowfly data) exhibit nonlinear features such as non-normality, nonlinearity between lagged variables and heteroscedasticity. They require nonlinear models to describe the law that generates the data. The most common nonlinear models are the threshold autoregressive (TAR) models [2], the exponential autoregressive (EXPAR) models [3], the smooth-transition autoregressive (STAR) models [4], the bilinear models [5], the random coefficient models [6], the autoregressive conditional heteroscedastic (ARCH) models [7]. However, nonlinear parametric modeling also has its drawbacks. Most importantly, it requires an a priori choice of parametric function classes for the function of interest. Thus, nonlinear parametric modeling implies the difficult choice of a model class. In contrast, when using the nonparametric modeling approach, one can avoid this choice.

Forecasting of the future values is one of the most popular application of time series modeling. In order to verify the accuracy of the forecast we need to define the error of prediction, which can be treated as a measure of uncertainty of the forecast. A close related problem is the construction of prediction intervals for future observations. In this purpose, for Gaussian data we use a well-known strategy. On the other hand, the Gaussian prediction intervals perform not very well for non-Gaussian series. In this context, an external bootstrap method will be proposed.

The bootstrap is a computer-intensive method that provides answers to a large class of statistical inference problems without stringent structural assumptions on the underlying random process generating the data. Since its introduction by Efron [**?**], the bootstrap has found its application to a number of statistical problems, including many standard ones, where it has outperformed the existing methodology as well as to many complex problems involving independent data where conventional approaches failed to provide satisfactory answers. However, the general perception that the bootstrap is an 'omnibus' method, giving accurate results in all problems automatically, is misleading. A prime example of this appears in Singh [**?**], which points out the inadequacy of this resampling scheme under dependence. A breakthrough was achieved with block resampling, an idea that was put forward by Hall [**?**] and others in various forms and in different inference problems. The most popular bootstrap methods for dependent data are block, sieve [**?**], local and external bootstrap [**?**].

The article is organized as follows.

2. Nonparametric Autoregressive Models

2.1. NAR Structure

Given a time series $\{Y_t, t = 1, ..., n\}$, in general, we can assume that

$$Y_t = g(X_t) + \nu(X_t)\, e_t \tag{1}$$

where $X_t = Y_{t-1}, ..., Y_{t-p}$, g and ν are unknown functions, and $\{e_t\} \sim IID\left(0, \sigma^2\right)$. Instead of imposing a concrete form on g and ν, we only make some qualitative assumptions, such as that the functions $g \in C^{\infty}(\mathbb{R})$ and $\nu \in C^{\infty}(\mathbb{R})$. Model (1) is called a nonparametric autoregressive conditional heteroscedastic (NARCH) model. The structure in (1) is very general, making very few assumptions on how the data were generated. It allows heteroscedasticity. In this paper we consider only a NAR structure ($\nu(\cdot)$ is a constant).

2.2. Classes of nonparametric estimators

In this subsection we review some nonparametric methods for estimating the function g in (1).

Model (1) has the format of a nonlinear regression problem for which many smoothing methods exists when the observations are independent. Hart [9] demonstrates that these methods can be "borrowed" for time series analysis where observations are correlated by making use of the "whitening by windowing principle".

The description concerning the four paradigms in nonparametric regression is based on [11]. The kernel estimate (local averaging) is due to [12], and [13]. The principle of least squares (global modeling) is much older. For historical details we refer to [14]. The principle of penalized modeling, in particular, smoothing splines, goes back to [15]. Final, The principle of complexity regularization is due to [16], in particular for least squares estimates, see [17].

Nadaraya-Watson kernel estimate A typical situation for an application to a time series $\{Y_t, t = 1, ..., n\}$ is that the regressor vector X consists of past time series values $X_t = (Y_{t-1}, ..., Y_{t-p})^T$. Let $K : \mathbb{R}^p \to \mathbb{R}_+$ be a function called the kernel function, and let $h > 0$ be a bandwidth. For $X \in \mathbb{R}^p$, $X_t = (Y_{t-1}, ..., Y_{t-p})^T$ and weights

$$w_{n,i}(x) = \frac{K\left(\frac{x - X_t}{h}\right)}{\sum_{i=p+1}^{n} K\left(\frac{x - X_i}{h}\right)} \tag{2}$$

where $w_{n,i} : \mathbb{R}^p \to \mathbb{R}$. The Nadaraya-Watson kernel estimator in model (1) with $\nu(\cdot)$ constant is given by

$$\hat{g}_n(x) = \sum_{i=p+1}^{n} w_{n,i}(x) Y_i. \tag{3}$$

For X equal to the last observed pattern, $X = (Y_n, Y_{n-1}, ..., Y_{n-p+1})^T$ this provides a one-step ahead predictor for Y_{n+1}. A k-step ahead predictor is given if Y_t in (3) is replaced by Y_{t-k+1}

$$\hat{g}_{n,k}(x) = \sum_{i=p+1}^{n} w_{n',i-k+1}(x) Y_i, \quad k = 1, 2, ... \tag{4}$$

where $n' = n - k + 1 - p$.

LS SVM regression

$$\hat{g}_{h,\gamma} = \sum \alpha_t K\left(\frac{x - X_t}{h}\right) + b \tag{5}$$

3. Construction of prediction intervals

Methods to establish confidence intervals or prediction intervals are based on the principle of first estimating $g(x)$ by an initial estimator $\hat{g}_h(x)$ and then estimating the distribution of $g(x) - \hat{g}_n(x)$. In the statistical literature a distinction is made between pivotal and nonpivotal methods. Hall [18] pointed out that pivotal methods, for the problem of bootstrap prediction intervals, should be preferred to nonpivotal methods.

Definition 1 (Pivotal quantities) *Let $X = (X_1, ..., X_n)$ be random variables with unknown joint distribution F, and let $T(F)$ denote a real-valued parameter. A random variable $\mathcal{J}(X, T(F))$ is a pivotal quantity (or pivot) if the distribution of $\mathcal{J}(X, T(F))$ is independent of all parameters. That is, if $X \sim F(x|T(F))$, then $\mathcal{J}(X, T(F))$ has the same distribution for all values of $T(F)$.*

Given a function estimator $\hat{g}_n(x)$, confidence intervals are constructed by using the asymptotic distribution of a pivot statistic. Let $\mathcal{J}(g(x), \hat{g}_n(x))$ be a pivotal statistic defined as

$$\mathcal{J}(g(x), \hat{g}_n(x)) = \frac{\hat{g}_n(x) - g(x)}{\sqrt{V(x)}}, \tag{6}$$

where $V(x)$ is the variance of the function estimator $\hat{g}_n(x)$. By following a procedure similar to that used for constructing confidence intervals for the function estimate, one can construct a prediction interval. In the pivot (6), one simply replaces the standard deviation of estimation $\sqrt{V(x)}$ by the standard deviation of prediction $\sqrt{\sigma^2(x) + V(x)}$.

3.1. Gaussian strategy

We first state the asymptotic distribution for the Nadaraya-Watson kernel estimator (3) and give the required assumptions. Let $f(x)$ denote the density of the lag vector at the point x. Then the asymptotic normal distribution for the Nadaraya-Watson kernel estimator (3) is given by

$$\sqrt{nh}\left(\hat{g}_n(x) - g(x)\right) \xrightarrow{d} \mathcal{N}(B(x), V(x)) \tag{7}$$

(see Härdle, 1990), where the bias is given by

$$B(x) = \frac{1}{2} c_0^{\frac{5}{2}} d_K f^{-1}(x) \left(g''(x) f(x) + 2g'(x) f'(x)\right), \tag{8}$$

and the variance is given by

$$V(x) = f^{-1}(x) c_K \sigma^2(x) \tag{9}$$

where $c_K = \int K^2(u)\, du$, $d_K = \int u^2 K(u)\, du$, c_0 is the constant such that $hn^{\frac{1}{5}}$ tends to c_0 in probability. Inspecting the asymptotic bias term (8) more closely reveals that the second-order derivatives of $g(x)$ have to exist. In fact, for (7) to hold this has to be the case in a neighborhood of x. For this reason one has to assume

that $g(x) \in C^2(\mathbb{R})$. Because both the density $f(x)$ and the conditional variance $\sigma^2(x)$ enter the asymptotic variance (9), one also has to assume that both are continuous and the latter is positive on the support of $f(x)$. For instance, we can estimate all the unknown terms in $B(x)$ and $V(x)$, depending on g and f, by using the kernel technique once more. A consistent bias estimate requires the estimation of second-order derivatives. Such estimates can be prone to a large variance, particularly if p is large and the sample size n is small. Thus, it make sense to compute prediction intervals without the bias correction.

¿From the asymptotic distribution (7), without the bias term, one can derive an asymptotic $(1-\alpha)$ percent prediction interval for $g(x)$,

$$P\left(\begin{array}{c}\hat{g}_n(x) - z_{\frac{\alpha}{2}}\sqrt{\sigma^2(x) + \frac{V(x)}{nh}} \leq g(x) \\ \leq \hat{g}_n(x) + z_{\frac{\alpha}{2}}\sqrt{\sigma^2(x) + \frac{V(x)}{nh}}\end{array}\right) = 1-\alpha \tag{10}$$

where $z_{\frac{\alpha}{2}}$ denotes the $\left(1-\frac{\alpha}{2}\right)$ quantile of the normal distribution.

3.2. Bootstrap strategy

3.2.1. Computing prediction intervals

We consider an alternative method that consists in estimating the distribution of the pivot

$$\mathcal{L}(g(x), \hat{g}_n(x)) = \frac{\hat{g}_n(x) - g(x)}{\sqrt{\sigma^2(x) + \hat{V}(x)}} \tag{11}$$

by an external bootstrap method. One approximate the distribution of the pivot statistics $\mathcal{L}(g(x), \hat{g}_n(x))$ by the corresponding distribution of the bootstrapped statistics

$$\mathcal{T}\left(\hat{g}_{n,h_2}(x), \hat{g}^*_{n,h_1}(x)\right) = \frac{\hat{g}^*_{n,h_1}(x) - \hat{g}_{n,h_2}(x)}{\sqrt{\sigma^2(x) + \hat{V}^*(q)}} \tag{12}$$

where $*$ denotes bootstrap counterparts. Given new input data, m simultaneous prediction intervals (applying the Bonferoni method) with asymptotic level $(1-\alpha)$ are given by

$$\mathcal{I}_T = \left(\begin{array}{c}\hat{g}_{n,h_1}(x) + \sqrt{\sigma^2(x) + \hat{V}^*(x)}Q_{\frac{\alpha}{2m}}, \\ \hat{g}_{n,h_1}(x) + \sqrt{\sigma^2(x) + \hat{V}^*(x)}Q_{\frac{(1-\alpha)}{2m}}\end{array}\right) \tag{13}$$

where Q_α denote the α-quantile of the bootstrap distribution of the pivotal statistic $\mathcal{T}(\hat{g}_g(x_0), \hat{g}^*_h(x_0))$.

3.2.2. External bootstrap

Algorithm 2 *Wild bootstrap.*

1. The unknown probability model P was taken to be $Y_t = g(Y_{t-1}, ..., Y_{t-p}) + e_t$, with $e_1, ..., e_n$ independent identically random error drawn from some unknown probability distribution function F_e.
2. Calculate $\hat{g}_n(X_t)$, and the estimated errors (residuals) are $\hat{e}_t = Y_t - \hat{g}_n(X_t)$, from which was obtained an estimated version of $\hat{F}_e$: probability $1/n$
3. Draw the bootstrap residuals $\hat{e}^*_t$ from a two-point centered distribution in order that its second and third moment fit the square and the cubic power of the residual $\hat{e}_t$. For instance, one can choose $\hat{e}^*_t \sim \hat{e}_t\left(\frac{Z_1\sqrt{2}}{+}Z_2^2 - 1\right)$, with Z_1 and Z_2 being two independent standard Normal random variables, also independent of $\hat{e}_t$.
4. Having generated Y^*_t, $t = 1, ..., n$, calculate the bootstrap estimates $\hat{g}^*_n(X_t)$.
5. This whole proces must be repeated B times.

The number of bootstrap replications, B, depend on the statistical inference problem. $B = 50$ is often enough to give a good estimate of a standard error. Much bigger values of B are required for bootstrap confidence (prediction) intervals. For example, $B = 1000$ [**?**]

4. Simulation

4.1. Toy example

4.2. Real data set

References

[1] Härdle, W. and Mammen, E., *Comparing nonparametric versus parametric regression fits*, Annals of Statistics, 21, No. 4, 1926-1947 (1988)

[2] Tong, H., *Nonlinear Time Series Analysis: A Dynamic Approach*, Oxford University Press, Oxford, (1990)

[3] Haggan, V. and Ozaki, T., *Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model*, Biometrika 68: 189-196, (1981)

[4] Chan, K.S. and Tong, H., *On estimating threshold in autoregressive models*, Journal of Time Series Analysis 7: 179-190, (1983)

[5] Granger, C.W.J. and Anderson, A.P., *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen and Zürich (1978)

[6] Nicholls, D.F. and Quinn, B.G., *Random Coefficient Autoregressive Models: An Introduction*, Vol. 11 of Lecture Notes in Statistics, Springer-Verlag, New York, (1982)

[7] Engle, R.F., *Autoregressive conditional geteroscedasticity with estimates of the variance of uk inflation*, Econometrica 50: 987-1008 (1982)

[8] Hart, D.J., *Nonparametric Smoothing and Lack-of-fit Tests*, Springer-Verlag, New York, (1997)

[9] Efron, B., *Bootstrap methods: Another look at the jackknife*, Annals of Statistics, 7: 1-26 (1979)

[10] Singh, K., *On the asymptotic accuracy of the Efron's bootstrap*, Annals of Statistics, 9: 1187-1195 (1981)

[11] Hall, P., *Resampling a coverage pattern*, Stochastic Processes and Their Applications 20, 231-246 (1985)

[12] Bühlmann, P., *Sieve bootstrap for time series*, Bernoulli 3, 123-148 (1997)

[13] Györfi, L., Kohler, M., Krzyżak, A., Walk, H., *A distribution-free theory of nonparametric regression*, Springer-Verlag, New York, (2002)

[14] Nadaraya, E.A., *On nonparametric estimates of density functions and regression curves*, Theory of Applied Probability 10, 186-190 (1965)

[15] Watson, G.S., *Smooth regression analysis*, Sankhya 26:15 175-184 (2002)

[16] Stigler, S.M., *Statistics on the table: The History of Statistical Concepts and MEthods*, Harvard University Press, Cambrigde, MA. (1999)

[17] Reinsch, C., *Smoothing by spline functions*, Numerische Mathematik, 10: 177-183 (1967)

[18] Vapnik, V.N., *Statistical Learning Theory*, John Wiley and Sons, INC. (1999)

[19] Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J. *Least Squares Support Vector Machines*, World Scientific, Singapore (2002)

[20] Hall, P., *On the bootstrap confidence intervals in nonparametric regression*, The Annals of Statistics, Vol. 20, No. 2, 695-711 (1992)

[21] Efron, B. and Tibshirani, R.J., *An Introduction to the bootstrap*, Chapman and Hall, London (1993)