Prediction Intervals for NAR Model Structures Using a Bootstrap Method

De Brabanter J.[*,**], Pelckmans K.[**], Suykens J.A.K.[**] and Vandewalle J.[**]

[*] Hogeschool KaHo Sint-Lieven (Associatie KULeuven), B-9000 Gent, Belgium
[**] K.U.Leuven, ESAT-SCD/SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Abstract— We consider the problem of constructing nonparametric prediction intervals for a NAR model structure. Our approach relies on the external bootstrap procedure [1]. This method is contrasted with a more traditional approach relying on the Gaussian strategy, showing improved results.

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 underlying dynamics that generate a given data set. Linear time series models are often the starting point for modeling time series.

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. Common nonlinear models are threshold autoregressive (TAR) models, exponential autoregressive (EXPAR) models, smooth-transition autoregressive (STAR) models, bilinear models, random coefficient models, autoregressive conditional heteroscedastic (ARCH) models, see e.g. [2]. 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 applications 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 closely related problem is the construction of prediction intervals for future observations. In this purpose, for Gaussian data one uses a well-known strategy. On the other hand, Gaussian prediction intervals perform not very well for non-Gaussian series. In this paper, an external bootstrap method will be proposed for this purpose.

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 [4], 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 in many complex problems involving independent data where conventional approaches failed to provide satisfactory answers. However, the generally perception that the bootstrap is a general applicable method, giving accurate results in all problems automatically, is misleading. An example of this appeared in Singh [5], 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 [6] and others in various forms and in different inference problems. The most popular bootstrap methods for dependent data are block, sieve [7], local and external bootstrap [1].

This paper is organized as follows. In Section 2 we describe nonlinear function estimation by LS-SVM and the Nadaraya-Watson kernel. In Section 3 we discuss methods of constructing predicting intervals (Gaussian strategy and bootstrap strategy). Section 4 reports results on an artificial data set.

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 $X_t = (Y_{t-1}, ..., Y_{t-p})^T$, g and ν are unknown functions, and $\{e_t\} \sim iid\left(0, \sigma^2\right)$. Instead of imposing a specific 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 (where $\nu(\cdot)$ 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 exist when the observations are independent. Hart [3] 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 kernel estimate (local averaging) is due to [8]. The principle of complexity regularization is due, e.g. to [9], in particular for least squares estimates, see [10].

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

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

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

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

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}

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

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

LS-SVM regression. Consider the model

$$g_{n,LS}(x) = w^T \varphi(x) + b \tag{5}$$

with so-called feature map $\varphi : \mathbb{R}^p \to \mathbb{R}^{D_\varphi}$, $w \in \mathbb{R}^{D_\varphi}$ and $b \in \mathbb{R}$. Consider the regularized least squares cost function [10]

$$\begin{aligned} &\min_{w,e,b} \mathcal{J} = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=p}^{n} e_t^2 \\ &\text{s.t.} \quad w^T \varphi(x_t) + b + e_t = y_t \quad \forall t = p+1, \ldots, n \end{aligned} \tag{6}$$

Then the dual solution is characterized by the following linear system

$$\left[\begin{array}{c|c} 0 & 1_n^T \\ \hline 1_n & \Omega + \frac{1}{\gamma} I_n \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline y \end{array}\right], \tag{7}$$

where $\Omega \in \mathbb{R}^{N \times N}$ with $\Omega_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, e.g. $\Omega_{ij} = K\left(\frac{x_i - x_j}{h}\right)$ for all $i, j = 1, \ldots, N$ and $y = (y_1, \ldots, y_n)^T$. The estimated model can be evaluated at a new point $x_* \in \mathbb{R}^p$ as follows

$$\hat{g}_{n,LS}(x) = \sum_{t=p+1}^{n} \alpha_t K\left(\frac{x - X_t}{h}\right) + b. \tag{8}$$

3. Construction of prediction intervals

The confidence (prediction) interval for nonparameteric regression falls into two parts, the first being the construction of a confidence (prediction) interval for the expected value of the estimator and the second involving bias correction. In the statistical literature a distinction is made between pivotal and nonpivotal methods. Hall [11] pointed out that pivotal methods, for the problem of bootstrap prediction intervals, should be preferred over 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 $\mathcal{R}(X, T(F))$ is a pivotal quantity (or pivot) if the distribution of $\mathcal{R}(X, T(F))$ is independent of all parameters. That is, if $X \sim F(x|T(F))$, then $\mathcal{R}(X, T(F))$ 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 $\mathcal{R}(g(x), \hat{g}_n(x))$ be a pivotal statistic defined as

$$\mathcal{R}(g(x), \hat{g}_n(x)) = \frac{\hat{g}_n(x) - g(x)}{\sqrt{V(x)}}, \tag{9}$$

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 (9), one simply replaces the standard deviation of $\sqrt{V(x)}$ by the standard deviation of prediction $\sqrt{\sigma^2(x) + V(x)}$. In the homoscedastic case $\nu = \sigma^2$ can be estimated (see [12] and references therein).

The effect of bias depends very much on how bias is corrected and there are different views amongst statisticians as how this should be done (explicit bias correction or undersmoothing techniques). The problem of

bias correction is described in the subsection "Gaussian strategy" and the subsection "Bootstrap strategy".

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, denoted by $\mathcal{N}$, for the Nadaraya-Watson kernel estimator (3) is given by

$$\sqrt{nh}\left(\hat{g}_n(x) - g(x)\right) \xrightarrow{d} \mathcal{N}\left(B(x), V(x)\right). \quad (10)$$

The symbol $\xrightarrow{d}$ defines equality in distribution. The bias of the estimator $\hat{g}_n$ is given by

$$B(x) = \frac{1}{2} c_0^{\frac{5}{2}} d_K f^{-1}(x) \left(g''(x) f(x) + 2 g'(x) f'(x)\right), \quad (11)$$

and the variance is given by

$$V(x) = f^{-1}(x) c_K \sigma^2(x) \quad (12)$$

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 (see [14]). Inspecting the asymptotic bias term (11) more closely reveals that the second-order derivatives of $g(x)$ have to exist. In fact, for (10) 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 (12), 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 may lead 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 (10), without the bias term, one can derive an asymptotic $(1-\alpha)$ percent prediction interval for $g(x)$,

$$P(\hat{g}_n(x) - z_{\frac{\alpha}{2}} \sqrt{\sigma^2(x) + \frac{V(x)}{nh}} \leq g(x) \leq \hat{g}_n(x) + z_{\frac{\alpha}{2}} \sqrt{\sigma^2(x) + \frac{V(x)}{nh}}) = 1 - \alpha \quad (13)$$

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

3.2. Bootstrap strategy

Based on the theorems of [1] and [11] we consider an alternative method that consists in estimating the distribution of the pivot

$$\mathcal{R}\left(g(x), \hat{g}_n(x)\right) = \frac{\hat{g}_n(x) - g(x)}{\sqrt{\sigma^2(x) + \hat{V}(x)}} \quad (14)$$

by an external bootstrap method. One approximates the distribution of the pivot statistics $\mathcal{R}\left(g(x), \hat{g}_n(x)\right)$ by the distribution of the bootstrapped statistics

$$\mathcal{T}\left(\hat{g}_{n,h_2}(x), \hat{g}^*_{n,h_1}(x)\right) = \frac{\hat{g}^*_{n,h_1}(x) - \hat{g}_{n,h_2}(x)}{\sqrt{\sigma^2(x) + \hat{V}^*(x)}} \quad (15)$$

where $*$ denotes bootstrap counterparts and h_1, h_2 are smoothing kernel parameters (a typical choice is $h_2 = ch_1$ with $c = 0.75$). Given new input data, m simultaneous prediction intervals (applying the Bonferoni method [15]) with asymptotic level $(1-\alpha)$ are given by

$$\mathcal{I}_T = [\hat{g}_{n,h_1}(x) + \sqrt{\sigma^2(x) + \hat{V}^*(x)} Q_{\frac{\alpha}{2k}}, \hat{g}_{n,h_1}(x) + \sqrt{\sigma^2(x) + \hat{V}^*(x)} Q_{\frac{(1-\alpha)}{2k}}] \quad (16)$$

where Q_α denote the α-quantile of the bootstrap distribution of the pivotal statistic $\mathcal{T}\left(\hat{g}_g(x_0), \hat{g}^*_h(x_0)\right)$. One has the following algorithm for the external bootstrap procedure.

Algorithm 1 *(External 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 errors drawn from some unknown probability distribution function F_e.*

2. *Calculate $\hat{g}_n(X_t)$. The estimated errors are $\hat{e}_t = Y_t - \hat{g}_n(X_t)$, from which one obtains an estimated version of $\hat{F}_e$ with probability $1/n$.*

3. *Draw the bootstrap residuals $\hat{e}^*_t$ from a two-point centered distribution such that its second and third moment fit the square and the cubic power of the residual $\hat{e}_t$. For instance, one can choose $\hat{e}^*_t \overset{i.i.d.}{\sim} \hat{e}_t \left(\frac{Z_1}{\sqrt{2}} + \frac{Z_2^2 - 1}{2}\right)$, with Z_1 and Z_2 being two independent standard Normal random variables (see [1]), also independent of $\hat{e}_t$.*

4. *Having generated $Y^*_t = \hat{g}_n(X_t) + e^*_t$ $t = 1, ..., n$, calculate the bootstrap estimates $\hat{g}^*_n(X_t)$.*

5. *This whole proces must be repeated for example $B = 1000$ times (see [13]).*

4. Illustrative example

To illustrate the prediction interval methods, we present an example using the following data set. Consider the nonlinear AR model defined on

$$Y_t = 3.7X_{t-1}(1 - X_{t-1}) + e_t, \quad t = 1, \ldots, 100 \quad (17)$$

where $e_t \overset{i.i.d.}{\sim} U[-0.1, 0.1]$. The logistic function, the

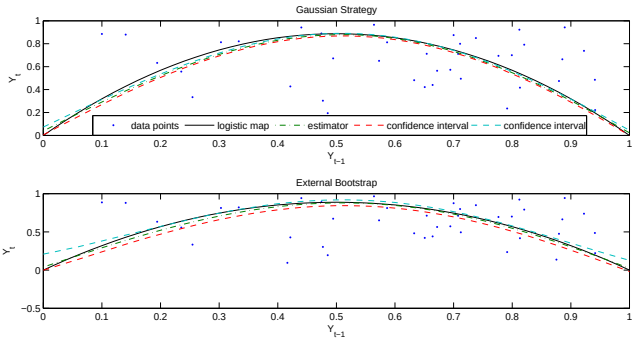

Figure 1: *The estimated regression function (dashdot lines) and its associated* 95% *confidence intervals (dashed lines).*

estimator and its associated 95% confidence intervals are given in Figure 1 for both strategies. Figure 2

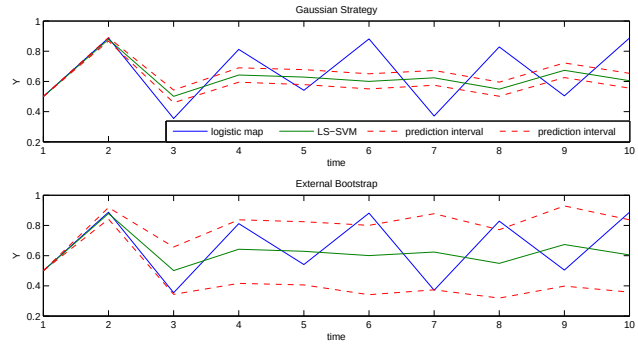

Figure 2: *The k-steps ahead,* $k = 1, \ldots, 10$*, predictor (non recurrent) and its associated* 95% *prediction intervals (dashed lines). The Gaussian strategy fails in this case. The bootstrap shows correct estimates.*

shows the improvements of the prediction intervals based on the bootstrap strategy in comparision with the prediction intervals based on the Gaussian strategy. Prediction intervals based on the bootstrap strategy enclose both the k-steps ahead predictor and the true underlying function.

Acknowledgements. This research work was carried out at the ESAT laboratory of KULeuven. Research Council KU Leuven: Concerted Research Action GOA-Mefisto 666, GOA-Ambiorics IDO, several PhD/postdoc & fellow grants; Flemish Government: Fund for Scientific Research Flanders (several PhD/postdoc grants, projects G.0407.02, G.0256.97, G.0115.01, G.0240.99, G.0197.02, G.0499.04, G.0211.05, G.0080.01, research communities ICCoS, ANMMM), AWI (Bil. Int. Collaboration Hungary/ Poland), IWT (Soft4s, STWW-Genprom, GBOU-McKnow, Eureka-Impact, Eureka-FLiTE, several PhD grants); Belgian Federal Government: DWTC IUAP IV-02 (1996-2001) and IUAP V-10-29 (2002-2006) (2002-2006), Program Sustainable Development PODO-II (CP/40); Direct contract research: Verhaert, Electrabel, Elia, Data4s, IPCOS. JS is an associate professor with K.U. Leuven.

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

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

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

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

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

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

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

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

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

[12] Pelckmans, K., J. De Brabanter, J.A.K. Suykens and B. De Moor, The Differogram: Nonparametric Noise Variance Estimation and its use for Model Selection, Special issue *Neurocomputing*, in press.

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

[14] Härdle, W., *Applied Nonparametric Regression*, Cambridge University Press (1990)

[15] Casella, G. and Berger, R.L., *Statistical Inference*, Duxbury Press, California (1990)