Quaderni di Statistica
Vol. 13, 2011

Value-at-Risk inference with NN-Sieve bootstrap

Francesco Giordano
Department of Economics and Statistics, University of Salerno
E-mail:giordano@unisa.it

Michele La Rocca
Department of Economics and Statistics, University of Salerno
E-mail:larocca@unisa.it

Cira Perna
Department of Economics and Statistics, University of Salerno
E-mail:perna@unisa.it

Summary: The aim of this paper is to construct bootstrap inference for VaR using a nonparametric bootstrap scheme, the NN-sieve bootstrap. This procedure, which retains the conceptual simplicity of the classical residual bootstrap, delivers consistent results for quite general nonlinear processes.
In this paper, we consider stochastic volatility models for financial time series of the nonlinear autoregressive-ARCH type and, in this context, we prove the consistency of the conditional quantile function estimator and we derive its asymptotic distribution. The proposed procedure is also evaluated through a small Monte Carlo study. The results confirm that the bootstrap quantile estimators converge, in some sense, to a Normal distribution. Moreover their distibutions are centered around zero and the variability decreases when the sample size increases, supporting the consistency of the procedure.

Keywords: Neural Networks, Bootstrap, VaR.

1. Introduction

In the last years the Value-at- Risk (VaR) concept has become one of the major tool in market risk management and a large amount of research has been dedicated to produce better VaR estimates. From a statistical point of view, VaR computation requires the estimation of a quantile of a return distribution. As soon as the probability distribution

of a return is specified, the VaR is calculated using the α-th quantile of this distribution. In this context, in addition to parametric VaR models, which require strong assumptions on the form of return distribution, several studies have investigated the possibility of using non parametric techniques. In this context, the literature suggests some interesting proposal; for example Cai and Wang (2008) use a combination of the weighted Nadaraya Watson estimator of Cai (2002) and the double kernel local linear estimation of Yu and Jones (1998). The advantage of this approach is especially that it is conceptually simple and easy to implement. Generally, the price to be paid for the flexibility of this tool is a slower convergence with respect to parametric models and, for fully non parametric models, the need of a large number of data points to estimate tail quantiles.
Alternatively, Franke and Diagne (2006) propose the use of feedforward neural networks for non parametric VaR estimation. One of the major benefit of this flexible approach is that it allows to take various types of intra and intermarket information into account.

The aim of this paper is to construct bootstrap inference for VaR using the Neural Network sieve bootstrap, a novel approach proposed by Giordano *et al.* (2005, 2009, 2011).
The basic idea of this proposal is to use feedforward neural network models as sieve approximators for nonlinear data generating processes. Several reasons justify this choice. First of all, ANNs provides an arbitrarily accurate approximation to unknown target functions which satisfy certain smoothness conditions. Barron (1993), and then Hornik et al. (1994) and Makovoz (1996), obtained a deterministic approximation rate (in L_2-norm) for a class of artificial neural networks with r hidden units and sigmoidal activation functions. Secondly, if the network model is fitted to the data in such a way that complexity of the network is allowed to increase at a proper rate with the sample size, the resulting function estimator can then be viewed as a nonparametric sieve estimator (Chen and Shen, 1998; Chen and White, 1999, Zhang 2004). Thirdly, estimation of hidden layer size seems to be less critical than estimation of the window size in local nonparametric approaches. Finally, artificial neural networks are global estimators and they do not suffer the so called 'curse of dimensionality': extension to high dimensional models is more straightforward than other nonparametric approaches.
The paper is organized as follows. In the next session the NN-sieve bootstrap is described. Some theoretical results are reported in section 3, while the results of a simulation experiment are discussed in section 4. Some remarks conclude the paper.

2. NN-Sieve Bootstrap

The basic idea of NN-sieve bootstrap is to use feedforward neural network models as sieve approximators.

Let $Y_t, t \in Z$ a stochastic process, modeled as

$$\begin{aligned} Y_t &= m\left(Y_{t-1}, \ldots, Y_{t-d}\right) + s\left(Y_{t-1}, \ldots, Y_{t-d}\right)\varepsilon_t \\ &= m\left(\mathbf{Z}_{t-1}\right) + s\left(\mathbf{Z}_{t-1}\right)\varepsilon_t \end{aligned} \tag{1}$$

where $m(.)$ and $s(.)$ are real valued functions defined on $\mathbb{R}^d$ and $\mathbf{Z}_{t-1} = (Y_{t-1}, \ldots, Y_{t-d})$. The errors $\{\epsilon_t\}$ are *i.i.d.* random variables with zero mean and unit variance. This model is useful to analyze financial time series characterized by nonlinear structures of the functions $m(.)$ and $s(.)$.
Of course, for $\mathbf{z} \in \mathbf{R}^d$, the function $m(\mathbf{z})$ represents the conditional mean function of the process while $s^2(\mathbf{z})$ is the volatility function, that is:

$$m\left(\mathbf{z}\right) = \mathbb{E}\left(Y_t | \mathbf{Z}_{t-1} = \mathbf{z}\right);$$

$$s^2\left(\mathbf{z}\right) = \text{var}\left(Y_t | \mathbf{Z}_{t-1} = \mathbf{z}\right).$$

The problem we focus on is the estimation of the conditional VaR which, following Franke and Diagne (2006), in this context is defined, as:

$$VaR(\mathbf{z}) = -q_\alpha(\mathbf{z})$$

where

$$q_\alpha(\mathbf{z}) = m\left(\mathbf{z}\right) + s\left(\mathbf{z}\right) q_\alpha^\varepsilon \tag{2}$$

and q_α^ε is the α-quantile of distribution of ε_t.

In order to estimate the quantity $q_\alpha(\mathbf{z})$ in (2), it is necessary to have an estimation of the functions $m(.)$, $s(.)$ and q_α^ε. The unknown functions $m(\cdot)$ and $s(\cdot)$ can be approximated by using single input, single layer feedforward neural network models in the class:

$$\mathcal{O}(r_n, \Delta_n) = \left\{ h_{r_n}\left(\mathbf{z}; \mathbf{w}\right) : \sum_{k=1}^{r_n} |c_k| < \Delta_n \right\}$$

with

$$h_{r_n}\left(\mathbf{z}; \mathbf{w}\right) = \sum_{k=1}^{r_n} c_k L\left(\mathbf{a}_k' \mathbf{z} + b_k\right) + c_0$$

where:

$L\left(\cdot\right)$ is a sigmoidal activation function;

$\mathbf{w} = (c_0, c_1, \ldots, c_{r_n}, \mathbf{a}_1, \ldots, \mathbf{a}_{r_n}, b_1, \ldots, b_{r_n})$;

$\{\mathbf{a}_k\}$ are the d dimensional vectors of weights for the connections between input layer and hidden layer;

$\{c_k\}$ are the weights of the link between the hidden layer and the output;

$\{b_k\}$, c_0 are, respectively, the bias terms of the hidden neurons and of the output and r_n is the hidden layer size.

A consistent estimate of the autoregression function $m(\cdot)$ can be obtained as

$$\hat{m} = \text{argmin}_{h \in \mathcal{O}(r_n, \Delta_n)} \frac{1}{n} \sum_{t=1}^{n} \left(Y_t - h(\mathbf{Z}_{t-1}; \mathbf{w})\right)^2.$$

Moreover, if $m_2(\mathbf{z}) = \mathbb{E}(Y_t^2 | \mathbf{Z}_{t-1} = \mathbf{z})$, an estimate of the squared of the volatility function s^2 can be obtained by:

$$\hat{s}^2(\mathbf{z}) = \hat{m}_2(\mathbf{z}) - \hat{m}(\mathbf{z})^2$$

where $\hat{m}_2$ is defined as

$$\hat{m}_2 = \text{argmin}_{h \in \mathcal{O}(r'_n, \Delta'_n)} \frac{1}{n} \sum_{t=1}^{n} \left(Y_t^2 - h(\mathbf{Z}_{t-1}; \mathbf{w}\prime)\right)^2.$$

If $\hat{q}_\alpha^\varepsilon$ is the α-quantile of an estimate of the distribution of ε_t, then by (2)

$$\hat{q}_\alpha(\mathbf{z}) = \hat{m}(\mathbf{z}) + \hat{s}(\mathbf{z})\, \hat{q}_\alpha^\varepsilon \tag{3}$$

estimates the conditional quantile function $q_\alpha(\mathbf{z})$. The consistency of this estimator and its asymptotic distribution is derived and discussed in the next session. In order to obtain the sampling distribution of the estimator (3), the following bootstrap procedure can be implemented.

First of all, the residuals from the network estimates

$$\hat{\varepsilon}_t = \frac{Y_t - \hat{m}(\mathbf{Z}_{t-1})}{\hat{s}(\mathbf{Z}_{t-1})} \tag{4}$$

are computed. Then the bootstrap replicates, which mimic the structure of the original series, are obtained from the recursion:

$$Y_t^* = \hat{m}\left(\mathbf{Z}_{t-1}^*\right) + \hat{s}\left(\mathbf{Z}_{t-1}^*\right) \varepsilon_t^*$$

where $\varepsilon_t^* \overset{iid}{\sim} F_{\hat{\varepsilon}}$, the empirical distribution function of the centered residuals with the first observation fixed to the mean value of Y_t and $t = 2, \ldots, n + n_1$. The first n_1 observations are discarded in order to make negligible the effect of starting values. The bootstrap VaR analogue is obtained by:

$$\widehat{VaR}^*(\mathbf{z}) = -\hat{q}_\alpha^*(\mathbf{z})$$

where:

$$\hat{q}_\alpha^*(\mathbf{z}) = \hat{m}(\mathbf{z}) + \hat{s}(\mathbf{z})\, \hat{q}_\alpha^{*\varepsilon} \tag{5}$$

and $\hat{q}_\alpha^{*\varepsilon}$ is the α-quantile of the correspondent bootstrap distribution.

As usual the bootstrap distribution can be approximated by Monte Carlo simulations. If the procedure is replicated B times, obtaining B bootstrap replicates of the statistic of interest $\hat{q}_{\alpha,b}^*(\mathbf{z})$, the empirical distribution function, conditionally on $\mathbf{Z}_{t-1} = \mathbf{z}$ is:

$$\hat{F}^* (x|\mathbf{z}) = B^{-1} \sum_{b=1}^{B} \mathbb{I} \left(\hat{q}^*_{\alpha,b}(\mathbf{z}) \leq x \right)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. It can be used to approximate the unknown sampling distribution of the estimator (3).

The proposed procedure has some advantages which makes it effective for the problem at hand (see also Giordano *et al.*, 2007). First of all, with respect the AR-sieve (Bühlmann 1997; 2002), the NN-sieve bootstrap is consistent for non linear processes and it exploits the good properties on NN modeling. Secondly, it does not exhibit artifacts in the dependence structure like in the blockwise bootstrap. Moreover, the NN-sieve bootstrap sample is not a subset of the original sample and there is no need of 'pre-vectorizing' the original observations. Thirdly, it enjoys the properties of a plug-in rule (for nonlinear data) and it retains the simplicity of the classical residual bootstrap while being a nonparametric bootstrap scheme. Finally, the NN-sieve bootstrap is shown to be asymptotically justified delivering consistent results for quite general non linear models, and it yields satisfactory results for finite sample size (Giordano *et al.*, 2011).

3. Some theoretical results

In order to formulate some theoretical results, it is necessary to assume that the data generating process is stationary, exponentially α-mixing, with the fourth moment finite as in Franke and Diagne (2006).

For sake of simplicity, in the following we refer to the model with $d = 1$; that is

$$Y_t = m\left(Y_{t-1}\right) + s\left(Y_{t-1}\right)\varepsilon_t. \tag{6}$$

However, the results can be easy extended to the general case for $d > 1$. For this model we assume that the following assumptions hold.

Assumptions **A**

A1) $\{\varepsilon_t\}$ is an *i.i.d.* sequence with a positive density function in every compact set in $\mathbb{R}$.

A2) $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = 1$, $E(\varepsilon_t^3) = 0$ and $E(\varepsilon_t^4) < \infty$.

A3) $|m(x)| \leq C_1(1+|x|)$ and $s(x) \leq C_2(1+|x|)$, with $C_1 \geq 0$, $C_2 \geq 0$ and $\forall x \in \mathbb{R}$.

A4) $\inf_{x\in\mathbb{S}} s(x) > 0$, for every compact set $\mathbb{S}$ in $\mathbb{R}$.

A5) $C_1 + C_2 \left(E(\varepsilon_t^4)\right)^{1/4} < 1$.

A6) $\{Y_t\}$ is a stationary process.

The assumptions (A) are sufficient conditions to guarantee that the process $\{Y_t\}$ is geometrically ergodic and exponentially β-mixing (Härdle and Tsybakov, 1997) and, as a consequence, also exponentially α-mixing.

We also need some conditions on the neural network structure. Let r_n and Δ_n be the number of neurons of the hidden layer and the upper bound for the weights of the Neural Network model, respectively. We suppose that the following assumptions hold.

Assumptions **B**

B1) $r_n \to \infty$ and $\Delta_n \to \infty$ when $n \to \infty$.

B2) The activation function, $L(\cdot)$, is sigmoidal and it is infinitely derivable.

B3) $(\Delta_n^2 r_n \log(\Delta_n^2 r_n))/\sqrt{n} \to 0$ when $n \to \infty$

Now, consider the estimator of the conditional quantile in (3) given that $Y_{t-1} = y$, as

$$\hat{q}_\alpha(y) = \hat{m}\left(y\right) + \hat{s}\left(y\right)\hat{q}_\alpha^\varepsilon \tag{7}$$

where $\hat{q}_\alpha^\varepsilon$ is the α-quantile estimator with respect to the empirical distribution function of $\hat{\varepsilon}_t$ as in (4), that is

$$\hat{\varepsilon}_t = \frac{Y_t - \hat{m}\left(Y_{t-1}\right)}{\hat{s}\left(Y_{t-1}\right)}. \tag{8}$$

Definition. We say that $X_n(y) \xrightarrow{P^*} X(y)$ if $P_Y\left(|X_n(y) - X(y)| > \epsilon\right) \xrightarrow{P} 0$ when $n \to \infty$, $\forall \epsilon > 0$.
Lemma 1. If the assumptions (A) and (B) hold then

$$\hat{q}_\alpha^\varepsilon \xrightarrow{P^*} q_\alpha^\varepsilon$$

where the conditions (B1) and (B3) hold for r_n, Δ_n and r'_n, Δ'_n with respect to the neural network estimators $\hat{m}(\cdot)$ and $\hat{m}_2(\cdot)$, respectively.

Proof.

By (8) and model (6), we can write, conditionally on $Y_{t-1} = y$

$$\hat{\varepsilon}_t | y = \varepsilon_t + \varepsilon_t\left(\frac{s(y)}{\hat{s}(y)} - 1\right) + \frac{s(y)}{\hat{s}(y)}\frac{\hat{m}(y) - m(y)}{s(y)} \tag{9}$$

Using the same arguments as in the proof of Corollary (4.1) in Franke and Diagne, (2006), it follows that

$$w_{tn}(y) := \varepsilon_t\left(\frac{s(y)}{\hat{s}(y)} - 1\right) + \frac{s(y)}{\hat{s}(y)}\frac{\hat{m}(y) - m(y)}{s(y)} \xrightarrow{P^*} 0 \quad \forall t, y$$

This result implies that $\hat{\varepsilon}_t|y \overset{P^*}{=} \varepsilon_t$ when $n \to \infty$. But, we can note that $\hat{\varepsilon}_t|y$, asymptotically, is independent of y.

By assumption (A2) and (A5) it implies that $E\left(Y_t^4\right) < \infty$. So we can conclude that $\sup_y w_{tn}(y) \xrightarrow{P^*} 0$. Since the random variables ε_t in $w_{tn}(y)$ do not depend on n and they are i.i.d. by assumption (A1), we can argue that there exists a n_0 such that $\forall n > n_0$, it follows that $w_n := sup_t sup_y w_{tn}(y) \xrightarrow{P^*} 0$, that is $\hat{\varepsilon}_t - \varepsilon_t \overset{P^*}{=} w_n$ when $n \to \infty$.

Now, let $F_{\hat{\varepsilon}}(u)$ be the empirical distribution function with respect to $\{\hat{\varepsilon_t}\}$. Then

$$F_{\hat{\varepsilon}}(u) = 1/n \sum_t \mathbb{I}\left(\hat{\varepsilon}_t \leq u\right)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

But $1/n \sum_t \mathbb{I}\left(\hat{\varepsilon}_t \leq u\right) = 1/n \sum_t \mathbb{I}\left(\varepsilon_t \leq u - w_n\right) = F_\varepsilon(u - w_n)$, where $F_\varepsilon(\cdot)$ is the empirical distribution function with respect to $\{\varepsilon_t\}$. By assumption (A1) the true distribution function of $\{\varepsilon_t\}$, $F(\cdot)$, is continuous, so we have that $F_\varepsilon(u - w_n)$ converges (in P^*) to $F(u)$, $\forall u$. Then, it follows that

$$F_{\hat{\varepsilon}}(u) - F(u) \xrightarrow{P^*} 0 \quad \forall u$$

Finally, applying Lemma (21.2) in van der Vaart, (1998), the result follows. □

Now we can consider the estimator $\hat{q}_\alpha(y)$ conditionally on $Y_{t-1} = y$.

Proposition 1. If the conditions in Lemma (1) hold then

$$\hat{q}_\alpha(y) - q_\alpha(y) \xrightarrow{P^*} 0$$

where $q_\alpha(y) = m\left(y\right) + s\left(y\right) q_\alpha^\varepsilon$.

Proof.

The proof is straightforward using Corollary (4.1) in Franke and Diagne, (2006) and Lemma (1). □

Remark. Using assumption (A1) it can be shown that $\sqrt{n}\left(\hat{q}_{n\alpha}^\varepsilon - q_\alpha^\varepsilon\right)$ converges in law to a Normal distribution with zero mean and variance $\alpha(1-\alpha)/f^2(q_\alpha^\varepsilon)$, where $\hat{q}_{n\alpha}^\varepsilon$ is the α-quantile estimator with respect to the empirical distribution function of $\{\varepsilon_t\}$ and $f(\cdot)$ is the density function of ε_t (van der Vaart, 1998, chapter 21).

4. Some simulation results

In this section we discuss the results of a small simulation experiment performed in order to evaluate the performance of the proposed procedure. In the data generating process we assume: $m(z) = 0$, $Y_t = s(Y_{t-1})\,\varepsilon_t$ with four different variance functions:

(M1) $s^2(z) = 0.7, \quad \varepsilon_t \sim N(0,1)$;
(M2) $s^2(z) = 0.1 + 0.3z^2, \quad \varepsilon_t \sim N(0,1)$;
(M3) $s^2(z) = 0.1 + 0.15z^2, \quad \varepsilon_t \sim T_{(10)}\sqrt{8/10}$;
(M4)$s^2(z) = 0.01 + 0.1z^2 + 0.35z^2\mathbb{I}(z < 0), \quad \varepsilon_t \sim N(0,1))$.

Model (M1) is clearly homoschedastic. Models (M2) and (M3) are ARCH models with, respectively, Gaussian and Student T innovation terms. The last one is scaled to be a unit variance random variable. Model (M4) is a threshold ARCH model introduced to evaluate the procedure in presence of asymmetric volatility function. In the Monte Carlo experiment, for each model we simulate $N = 300$, series of length $n = 500, 1000, 2000$. For each model, we generate $B = 1000$ bootstrap replicates. Feedforward neural network models have been estimated by using nonlinear least squares. The hidden layer size has been estimated by using the correct Akaike Information criterion, in order to keep as low as possible the complexity of the network.

To make inference about the conditional VaR, we consider two cases. First, we suppose that the bootstrap estimator $\hat{q}^*_\alpha(\mathbf{z})$, in (5), has an asymptotic Normal distribution. So, we have only to verify that the bootstrap variance of $\hat{q}^{*\varepsilon}_\alpha$ is consistent with respect to the true one, since the estimators of $m(\cdot)$ and $s(\cdot)$ are consistent as in Franke and Diagne, (2006).

In the second case we can use $\hat{q}^*_\alpha(\mathbf{z})$ to estimate the sample distribution function for the conditional VaR. In this context we have to verify that the bootstrap distribution of $\hat{q}^{*\varepsilon}_\alpha$ converges in some sense to a Normal distribution.

First, we use the above simulation experiment to verify that $nVar^*\left(\hat{q}^{*\varepsilon}_\alpha\right)$ converges in some sense to $nVar\left(\hat{q}^{\varepsilon}_{n\alpha}\right)$.

In figures (1) and (2) we can observe that the bootstrap estimator of $nVar\left(\hat{q}^{\varepsilon}_{n\alpha}\right)$ has a good performance for all the four models. In fact, the reference line (the true value) crosses the boxplot graphs in correspondence of the median values.

To analyze the sample distribution for the estimator of q^ε_α we consider $\gamma_1 = 0.01$ and $\gamma_2 = 0.05$. Let $z_{(\gamma_i)}, \quad i = 1, 2$, be the quantiles from a standard Normal distribution at levels γ_i. So we define

$$\hat{c}^*_{Normal}(\gamma_i) = \hat{q}^\varepsilon_\alpha + \sqrt{Var^*\left(\hat{q}^{*\varepsilon}_\alpha\right)}z_{(\gamma_i)} \quad i = 1, 2$$

as the γ_i quantiles using the asymptotic Normal distribution and $\hat{c}^*_{Boot}(\gamma_i)$ as the quantiles from the bootstrap distribution estimators. Both, $\hat{c}^*_{Normal}(\gamma_i)$ and $\hat{c}^*_{Boot}(\gamma_i)$, refer to the estimator $\hat{q}^\varepsilon_\alpha$ for the true quantiles $c(\gamma_i) = q^\varepsilon_\alpha + \sqrt{Var\left(\hat{q}^\varepsilon_\alpha\right)}z_{(\gamma_i)}$.

To compare these two methods for the estimation of $c(\gamma_i)$, we consider the statistics:

$$S_i = \frac{\hat{c}^*_{Normal}(\gamma_i) - \hat{c}^*_{Boot}(\gamma_i)}{c(\gamma_i)}, \quad i = 1, 2$$

in which the denominator allows to avoid the influence of n, the length of the time series.

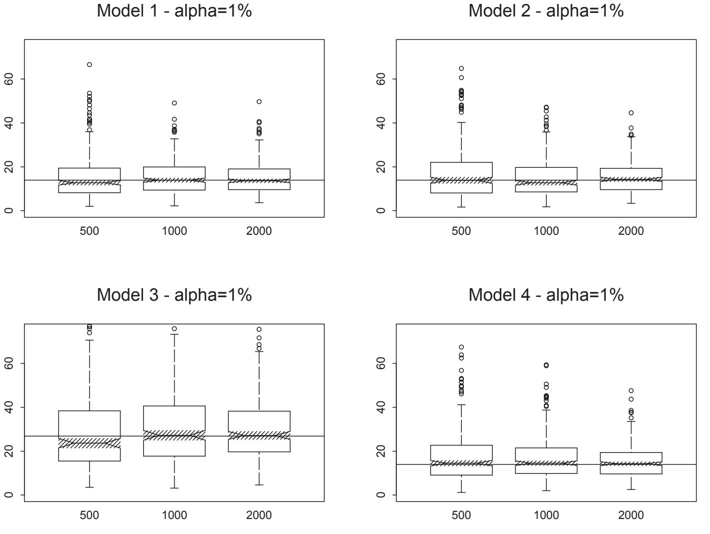

Figure 1. Distribution of $nVar^*\left(\hat{q}_{\alpha}^{*\varepsilon}\right)$*,* $\alpha = 0.01$*. Reference line is the true value* $nVar\left(\hat{q}_{n\alpha}^{\varepsilon}\right)$.

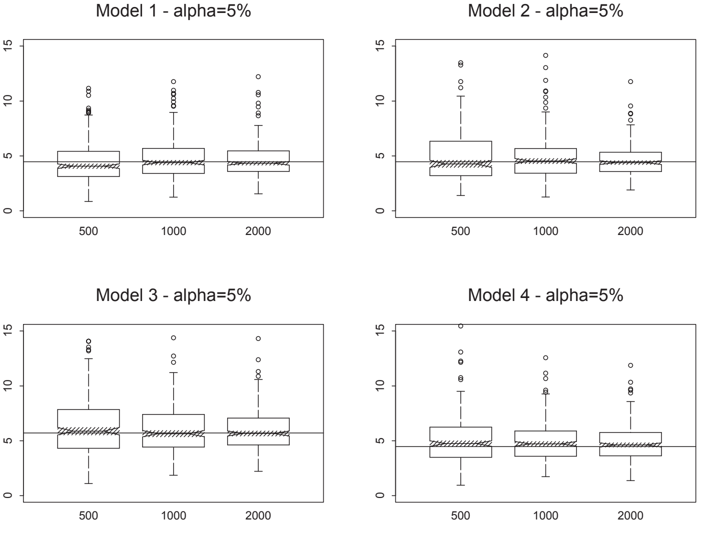

Figure 2. Distribution of $nVar^*\left(\hat{q}_{\alpha}^{*\varepsilon}\right)$*,* $\alpha = 0.05$*. Reference line is the true value* $nVar\left(\hat{q}_{n\alpha}^{\varepsilon}\right)$.

From figures (3) and (4), the bootstrap distribution of the estimators for q_α^ε seem to be well approximated by the Normal distribution. Moreover, the boxplot graphs seem to be centerd about the vales zero for all the time series lenghts whereas the variability decreases when n grows.

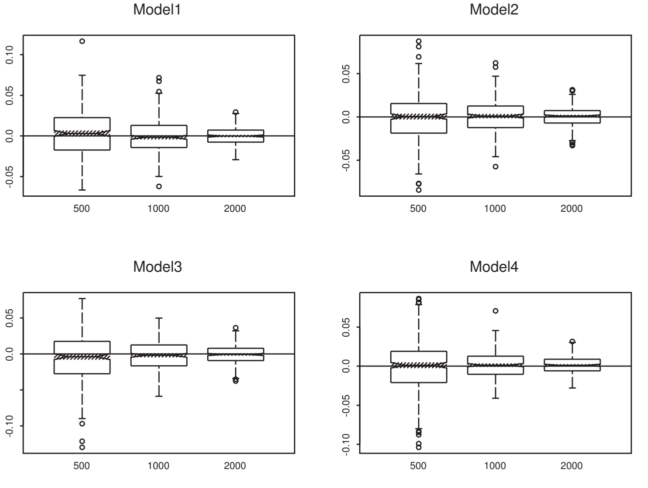

Figure 3. Distribution of S_1 with $\gamma = 0.01$, $\alpha = 0.05$. Reference line is at zero.

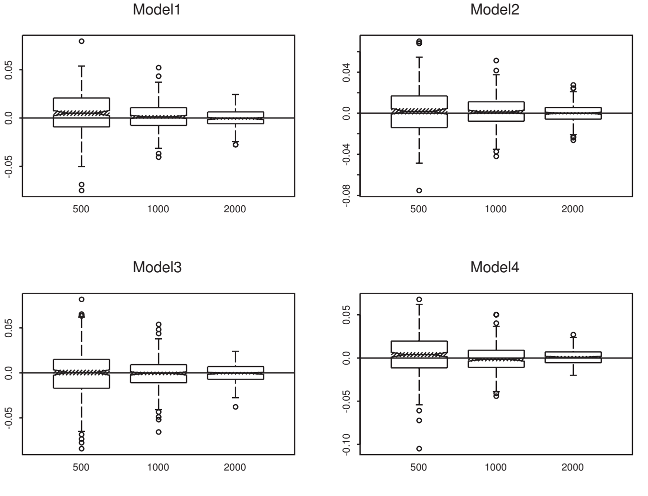

Figure 4. Distribution of S_2 with $\gamma = 0.05$, $\alpha = 0.05$, Reference line is at zero.

5. Concluding remarks

In this paper we have proposed the NN-sieve bootstrap for VaR inference. The basic idea of this proposal is to use feedforward neural network models as sieve approximators for nonlinear data generating processes. The approach, which is non-parametric in its spirit, retains the conceptual simplicity of a classical residual bootstrap and it does not have the problems of other nonparametric bootstrap techniques such as blockwise schemes. Moreover, it is shown to be asymptotically justified and it delivers consistent results for quite general nonlinear processes. We have considered stochastic volatility models for financial time series of the nonlinear autoregressive-ARCH type and, in this context, we have proved the consistency of the conditional quantile function estimator. We have also derived its asymptotic distribution.

The performances of the proposed procedure has been also disscussed by mean of small a Monte Carlo study. The results confirm that the bootstrap quantile estimators converge, in some sense, to a Normal distribution. Moreover their distibutions are centered around zero and the variability decreases when the sample size increases, supporting the consistency of the procedure.

Acknowledgements: The authors gratefully acknowledge support from the University of Salerno grant program "Sistema di calcolo ad alte prestazioni per l'analisi economica, finanziaria e statistica (High Performance Computing - HPC)- prot. ASSA098434, 2009

References

Barron A.R. (1993), Universal approximation bounds for superposition of a sigmoidal function, *IEEE Transactions on Information Theory*, 39, 930–945.

Bühlmann P. (1997), Sieve bootstrap for time series, *Bernoulli*, 3, 123–148.

Bühlmann P. (2002), Sieve bootstrap with variable-length Markov chains for stationary categorial time series, *Journal of the American Statistical Association*, 97, 443–471.

Cai Z. (2002), Regression quantiles for time series, *Econometric Theory*, 18, 169–192.

Cai Z., X. Wang (2008), Nonparametric estimation of conditional var and expected shortfall, *Journal of Econometrics*, 147, 120–130.

Chang Y., Park J. (2003), A Sieve Bootstrap for the Test of a Unit Root, *Journal of Time Series Analysis*, 24, 379–400.

Chen X., Shen X. (1998), Asymptotic properties of sieve extremum estimates for weakly dependent data with applications, *Econometrica*, 66, 299–315.

Chen X., White H. (1999), Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators, *IEEE Transactions on Information Theory*, 45, 682–691.

Franke J., Diagne M. (2006), Estimating market risk with neural networks, *Statistics & Decisions*, 24, 233–253.

Giordano F., La Rocca M., Perna C. (2005), Neural network sieve bootstrap for nonlinear time series, *Neural Network World*, 15, 327–334.

Giordano F., La Rocca M., Perna C. (2007), Forecasting nonlinear time series with neural network sieve bootstrap, *Computational Statistics & Data Analysis*, 51, 3871–3884.

Giordano F., La Rocca M., Perna C. (2009) Neural Network Sieve Bootstrap Prediction Intervals: Some Real Data Evidence. In B. Apolloini, S. Bassis, M. Marinaro *New Directions in Neural Networks: Frontiers in Artificial Intelligence and Applications*, 205–213.

Giordano F., La Rocca M., Perna C. (2011), Properties of the neural network sieve bootstrap, *Journal of Nonparametric Statistics* , 23, 803–817.

Hornik K., Stinchcombe M., Auer P. (1994), Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives, *Neural Computation*, 6, 1262–1275.

Härdle W., Tsybakov A. (1997), Local Polynomial estimators of the volatility function in nonparametric autoregression, *Journal of Econometrics*, 81, 223–242.

Makovoz Y. (1996), Random approximates and neural networks, *Journal of Approximation Theory*, 85, 98–109.

van der Vaart A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Yu K., Jones, M. C. (1998), Local linear quantile smoothing, *J. Amer. Statist. Assoc.*, 93, 228–238.

Zhang J. (2004), Sieve Estimates via Neural Network for Strong Mixing Processes, *Statistical Infrence for Stochastic Processes*, 7, 115–135.