ISSN 1440-771X

MONASH UNIVERSITY

AUSTRALIA

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Nonlinear Correlograms and Partial Autocorrelograms

Heather M. Anderson and Farshid Vahid

Working Paper 19/2003

Nonlinear Correlograms and Partial Autocorrelograms

Heather M. Anderson* and Farshid Vahid
Department of Econometrics
and Business Statistics
Monash University
Clayton, Victoria, 3800
AUSTRALIA

November 19, 2003

Abstract

This paper proposes neural network based measures of predictability in conditional mean, and then uses them to construct nonlinear analogues to autocorrelograms and partial autocorrelograms. In contrast to other measures of nonlinear dependence that rely on nonparametric estimation of densities or multivariate integration, our autocorrelograms are simple to calculate and appear to work well in relatively small samples.

Keywords: Nonlinear autocorrelograms, Nonlinear time series models, Neural networks, Model selection criteria, Nonlinear partial autocorrelograms
JEL classification: C22, C45, C51

*Correspondence: E-mail: Heather.Anderson@BusEco.monash.edu.au

This work was supported by the Australian Research Council (Large Grant No A00103258). We acknowledge helpful comments from Clive Granger, David Hendry, Bernt Stigum, Herman van Dijk, Kenneth Wallis, Andy Weiss and participants of the Econometric Methodology Conference held in Oslo, 2003. The usual disclaimer applies.

1 Introduction

This paper addresses lag selection in the modelling of the conditional mean of a nonlinear time series process. There are large literatures on testing for nonlinearity (see e.g. Ramsey (1969), Keenan (1985), Lee et al (1993), Teräsvirta et al (1994) and Brock et al (1996)), testing for neglected nonlinearity in an estimated model with prespecified alternatives (see, e.g. Tsay (1989), Hansen (1992) or Teräsvirta (1994)), and consistent tests of correct specification without specifying alternative models (see, e.g. Bierens (1990) and Hong and White (1995)). However, when the objective is to model a time series that is believed (perhaps by pre-testing) to be nonlinear, these tests do not suggest how many lags should enter the specification. Even though the modeller may have only one class of nonlinear models in mind, it would be impractical to suggest selecting the lag length using a usual likelihood based model selection criterion or a "general to simple" methodology, because estimating entire sets of nonlinear models is both time consuming and cumbersome.

The purpose here, is to suggest easy to compute modelling aids for the nonlinear specification of the conditional mean of a time series, given its past. These aids are similar to the autocorrelogram and the partial autocorrelogram that are used in linear contexts, but they are designed to detect lag structures that standard correlograms cannot find. They detect linear correlations as well, and therefore complement the information provided by standard correlograms, in both linear and nonlinear contexts. Our nonlinear autocorrelograms and partial autocorrelograms are based on conditional moment test statistics, and for the variable y_t we ask what is the longest lag p such that $E(y_t|y_{t-p})$ is a non-constant function of y_{t-p}. We also ask what is the longest lag p such that $E(y_t|y_{t-1},\ldots,y_{t-p+1}) \neq E(y_t|y_{t-1},\ldots,y_{t-p})$. In developing our correlation measures, we use results on using neural networks to approximate functions, and recent results on forecast combinations and common factors.

There is a small, but growing literature on lag dependencies in nonlinear contexts. This includes contributions by Auestad and Tjøstheim (1990), Tjøstheim and Auestad (1994) and Granger and Lin (1994). More recent developments include the Hong and White (2000) entropy measures of serial dependence, work by Gourieroux and Jasiak (2002), who have defined nonlinear canonical correlations between y_t and y_{t-p} as the maximal correlation between $g\left(y_t\right)$ and $E\left(g\left(y_t\right)|y_{t-p}\right)$ over all g, and work by Hong and Lee (2003), who develop tests of independence of y_t and y_{t-p} based on the generalized spectrum. Granger et al. (2003) have suggested a new measure of dependence between y_t and y_{t-p} based on the distance between the joint and product of marginal densities.

All of these new statistics are sometimes used as "nonlinear correlograms", but they focus on the dependence between y_t and y_{t-p}, rather than on whether y_{t-p} can be used to predict y_t. Obviously if y_t and y_{t-p} are independent, then y_t is not predictable from y_{t-p}. However, a lack of independence between y_t and y_{t-p} does not necessarily imply that y_t is predictable from y_{t-p}. The first moment of y_t can be independent of y_{t-p} while higher moments depend on y_{t-p}. Also, establishing that y_t and y_{t-p} are dependent does not by itself imply that $E(y_t \mid \mathcal{I}_{t-1})$ should include y_{t-p}[1]. These considerations lead to the conclusion that for lag selection, a nonlinear analogue of a partial autocorrelogram is needed in addition to dependence measures, and in our opinion none of the dependence measures reviewed above lend themselves easily to such a generalization.

When modelling nonlinear processes, it is common practice to use information criteria such as those proposed by Akaike (1974), Hannan and Quinn (1979) or Schwartz (1978) to select the lag length of a linear specification, and then to develop a nonlinear specification conditional on the chosen lag length. However, it is easy to imagine that such criteria might favour short lag structures when applied to some nonlinear data generating processes, especially if the nonlinear structure comes into play at relatively distant lags. In such cases, tests for nonlinearity based on the chosen linear null model might fail to find evidence of nonlinearity. At the other extreme and perhaps more importantly, the choice of too many lags for a linear approximation of a nonlinear DGP can imply many additional parameters in the estimated nonlinear model, leading to a highly overparameterized model that delivers poor forecasts. Anderson (2002) finds that all of the criteria cited above (i.e. AIC, HQ and BIC) tend to overpredict lag length when applied to nonlinear DGPs. She also finds that the selection of lag length by applying AIC, HQ and/or BIC to approximating quadratic models of the data ameliorates this overprediction, although the benefits of doing this are not substantial when samples are small.

Neural network models are very common in the nonlinearity literature, because they are relatively simple to use and can approximate most forms of nonlinearity well. Section 2 describes how we adapt and use these models to define measures of predictability of y_t from y_{t-p}. We use these measures to form nonlinear autocorrelograms, and then discuss the testing of whether or these 'correlations' are statistically different from zero. In Section 3, we use similar techniques to measure the predictability of y_t from y_{t-p} after accounting for $y_{t-1}, y_{t-2}, \ldots, y_{t-p+1}$. We use these measures to form nonlinear partial

[1] A nonlinear autoregressive process of order one can imply that y_t and y_{t-p} are dependent for $p > 1$, but in this case the correct specification for y_t does not include y_{t-p}.

autocorrelograms, and then discuss the testing of whether or these ‘partial correlations’ are statistically different from zero. Section 4 studies the empirical performance of these measures for a selection of linear and nonlinear DGPs. Our sample DGPs are relatively small (100 to 200) observations, so that our conclusions relate to the sorts of samples that econometricians typically encounter. After establishing that our nonlinear autocorrelograms and partial autocorrelograms work quite well for known DGPs, we then analyze some actual data that is known to exhibit nonlinear behaviour. Section 5 summarizes and concludes.

2 Nonlinear Autocorrelograms

Our nonlinear autocorrelograms are based on the fact that if $E\left(y_t \mid y_{t-p}\right) = 0$[2], then $Corr\left(y_t, f\left(y_{t-p}\right)\right) = 0$ for all measurable functions f, and if $E\left(y_t \mid y_{t-p}\right) = \Psi(y_{t-p})$ where Ψ is a non-trivial function, then we can find sufficiently many measurable functions f such that $Corr\left(y_t, f\left(y_{t-p}\right)\right) \neq 0$. Hornik et al (1989) show that we can approximate any nonlinear function Ψ of y_{t-p} arbitrarily well, by using a linear combination of q elementary functions of y_{t-p} for q sufficiently large. The approximating model of $\Psi(y_{t-p})$ is given by

$$E\left(y_t \mid y_{t-p}\right) \equiv \Psi(y_{t-p}) \simeq \alpha_0 + \sum_{j=1} \alpha_j \phi\left(\gamma_j'(1, y_{t-p})\right) \tag{1}$$

where ϕ is a permissible elementary function[3], and the γ_j are randomly chosen by the econometrician, independently of y_{t-p}. A large q allows y_{t-p} to influence $\Psi(y_{t-p})$ in many different directions, and the α_j weight these influences so that the aggregated sum of the $\alpha_j\phi\left(\gamma_j'(1, y_{t-p})\right)$ can approximate the nonlinearity very well.[4]

Equation (1) requires q to be “large” (i.e. approaching infinity) if the approximation is to work well, and this requirement can be prohibitive when working with small samples. Here, we obtain a set of m very crude approximations to $\Psi(y_{t-p})$ by setting

[2] Strictly speaking, we need $E\left(y_t \mid y_{t-p}\right)$ to be constant, but from now on we assume that y_t has been demeaned without loss of generality.

[3] The elementary function, which is called the “activation function” or the “squashing function” in the neural network literature, can be any function that satisfies some continuity and denseness conditions discussed in Hornik et al (1989). We use the logistic function $\phi(z) = [1 + \exp(z)]^{-1}$in this paper.

[4] For neural network modelling, one chooses the number of nodes judiciously and estimates the parameter of the q activation functions using a nonlinear optimization algorithm. Here, our goal is not to provide a neural network model for the time series (as in, say, Perez-Amaral et al (2003)), but rather to use neural network approximations of $\Psi(y_{t-p})$ to provide easy-to-calculate measures of nonlinear predictability.

$q = 1$ for each of m versions of (1) (each based on a single draw of γ_j') and then average the information obtained from this set of m predictions to obtain a "combined prediction" of $\Psi(y_{t-p})$. The rationale for this is based on the well known observation (see e.g. Granger (1989)) that forecast combinations based on many different predictor sets often work better than a single prediction based on a forecasting model that includes all predictors, so that an average taken over many simple network approximations of a given nonlinear model can account for complicated nonlinearities that might not be well captured by a single highly parameterized network. Recent work (see, e.g. Stock and Watson (2000), Elliott and Timmerman (2002), Hendry and Clements (2002) and Granger and Jeon (2002)) has studied various forecast combinations (such as simple or weighted averages of the forecasts) in a variety of settings, and has shown that forecast combinations can deliver very good predictions, even if none of the individual forecasting equations perform very well.

The first measure of non-linear autocorrelation at lag p that we propose is

$$nlac_1^0(p) = \frac{1}{m}\sum_{j=1}^{m} r^2\left(y_t, \phi\left(\gamma_j'(1, y_{t-p})\right)\right) \tag{2}$$

where each ϕ is based on just one random draw of γ_j and $r^2(.,.)$ is the square of sample correlation coefficient of its arguments. When y_t is unpredictable from y_{t-p}, then $E\left[y_t\phi\left(\gamma_j'(1, y_{t-p})\right)\right] = 0$ for all γ_j and therefore $nlac_1^0(p) \to 0$ in probability as T goes to infinity for any m. When $E(y_t \mid y_{t-p})$ is a non-trivial function of y_{t-p}, then $nlac_1^0(p)$ converges to a positive limit as m and T go to infinity. Moreover, when y_t is white noise, then $Tr^2(y_t, \phi\left(\gamma_j'(1, y_{t-p})\right))$ is asymptotically χ_1^2 for each j and any $p > 0$, so that the average over all j will be close to $E\left(\chi_1^2\right) = 1$. Therefore, as a rough guide, if $nlac_1^0(p) > \frac{1}{T} + (2 \times \frac{\sqrt{2}}{T})$ then there is strong evidence that lag p has predictive ability[5]. Our simulations provide better critical values than this conservative bound. In the case where y_t is unpredictable from y_{t-p}, but there is some non-trivial linear correlation between y_t and y_{t-s} for $s < p$, then $Tr^2(y_t, \phi\left(\gamma_j'(1, y_{t-p})\right))$ is only a χ_1^2 after it has been readjusted by the product of the ratios of the variance to the long-run variance of y_t and $\phi\left(\gamma_j'(1, y_{t-p})\right)$. However, as is usual practice with linear autocorrelograms, we ignore this subtlety and use $\frac{1}{T} + (2 \times \frac{\sqrt{2}}{T})$ as the critical bound.

Some may be concerned that our proposed measure of non-linear squared autocorrelation does not return a value of 1 when $p = 0$, i.e. when the non-linear correlation of y_t with itself is considered. This can be easily accommodated by considering a slightly

[5] Note that we have not divided the standard deviation by m. This is because the r^2s are not independent.

modified version of $nlac_1^0(p)$ given by

$$nlac_1(p) = \frac{1}{m}\sum_{j=1} R^2\left(y_t \text{ on } 1, y_{t-p}, \phi\left(\gamma_j'(1, y_{t-p})\right)\right), \tag{3}$$

in which R^2 is the usual regression R-squared. In this case, for a white noise process, TR^2 will have an asymptotic χ_2^2 distribution, and therefore the guideline critical bound is $\frac{2}{T} + (2 \times \frac{2}{T})$.

Both $nlac_1^0$ and $nlac_1$ are based on averages of predictability measures along random nonlinear directions. Another approach would be to attempt to find a "best" nonlinear direction first, and then to measure predictability along that dimension only. The requirement for the first stage is that it must be achieved through an algorithm that is guaranteed to work (in finite time) without human intervention. This excludes estimating the best neural network model using maximum likelihood, but it allows one to consider many 'single-draw' networks. Based on this idea, we propose the following,

$$nlac_2(p) = r^2(y_t, \text{weighted average of projections of } y_t \text{ onto } \mathcal{S}_j(y_{t-p})), \tag{4}$$

where $\mathcal{S}_j(y_{t-p})$ is the span of $1, y_{t-p}, \phi\left(\gamma_j'(1, y_{t-p})\right)$ for a randomly chosen γ_j. Again, we include $1, y_{t-p}$ to make sure that $nlac_2(0) = 1$, and we consider approximating models of the form of equation (1) with $q = 1$.[6] It is reasonable to expect that model averaging will give a better prediction than any of the individual models, and we weight each prediction by its respective R-squared. Using the direction with maximum fit would be dangerous, because this might simply fit the noise in y_t. Also, using a simple average of predictions would not be entirely satisfactory, because giving the same weight to directions with poor predictability as those with good predictability would introduce unnecessary noise. Since all of these predictions are predictions of y_t, this weighting scheme is the same as weighting by the variance of the predictions.

Our weighting scheme is also justified by thinking about the "best" predictor of y_t as the common factor of projections of y_t on many random directions parameterized by γ_j. A legitimate estimator of the common factor is the average of projections. However, if some of the projections carry only faint signals about the common factor, it is best to exclude them or weight them less than others (see Boivin and Ng 2002). This justifies a weighting scheme that reflects the strength of the signal.

The $nlac_2(p)$ measure is asymptotically justified if the number of nonlinear functions that span $\mathcal{S}_j(y_{t-p})$ is allowed to grow to infinity at an appropriate rate[7] that is

[6] We could also define $nlac_2^0(p) = r^2(y_t, \text{weighted average of projections of } y_t \text{ onto } \mathcal{S}_j(y_{t-p}))$, where $\mathcal{S}_j(y_{t-p})$ is the span of $\{1, \phi\left(\gamma_j'(1, y_{t-p})\right)\}$ for randomly chosen γ_j.

[7] See for example Chen and White (1999).

slower than T. Under this condition, each of the projections converge to the conditional expectation function, and so does the weighted average of them. However, the asymptotic justification provides no direction for how many nonlinear functions to include for a particular finite sample size T. Here, we have defined $\mathcal{S}_j\left(y_{t-p}\right)$ with only one nonlinear function of y_{t-p}, but of course it is desirable to use more than one if we are working with large samples. It is difficult to determine a rough critical value for this measure even when y_t is white noise because the weighting scheme biases the average towards directions with higher predictability. The critical value is therefore determined by simulation. The empirical performance of $nlac_1^0(p)$, $nlac_1(p)$, $nlac_2^0\left(p\right)$ and $nlac_2\left(p\right)$ is studied in Section 4.

3 Nonlinear Partial Autocorrelograms

As mentioned in the introduction, none of the measures of dependence suggested in the literature lend themselves easily to a partial measure of predictability. Our objective is to suggest measures of nonlinear partial autocorrelation that can be used for order selection in linear or nonlinear autoregressive models. Nonlinear partial autocorrelation of order p is a measure of predictability of $y_t - E\left(y_t|y_{t-1},...,y_{t-p+1}\right)$ from $\left(y_{t-1},...,y_{t-p+1},y_{t-p}\right)$. The partial autocorrelation must be zero when

$$E\left[\left(y_t - E\left(y_t \mid y_{t-1},...,y_{t-p+1}\right)\right) \mid \left(y_{t-1},...,y_{t-p+1},y_{t-p}\right)\right] = 0 \tag{5}$$

which implies that

$$E\left[\psi\left(y_{t-p}\right)\left(y_t - E\left(y_t \mid y_{t-1},...,y_{t-p+1}\right)\right)\right] = 0 \tag{6}$$

for all measurable functions ψ, and also that

$$E\left(y_t \mid y_{t-1},...,y_{t-p+1},y_{t-p}\right) = E\left(y_t \mid y_{t-1},...,y_{t-p+1}\right). \tag{7}$$

We base our measures of partial autocorrelation on equations (6) and (7).

The first measure quantifies the importance of adding random functions of y_{t-p} in regressions with y_t as the dependent variable and random functions of $y_{t-1},...,y_{t-p+1}$ as independent variables. Specifically, we draw p random numbers from appropriate distributions[8], arrange them into a vector γ_j and form $\phi \quad \gamma_j'(1,y_{t-1},...,y_{t-p+1})$, where ϕ is a logistic function. We run a regression of y_t on $1,y_{t-1},...,y_{t-p+1},\phi \quad \gamma_j'(1,y_{t-1},...,y_{t-p+1})$

[8] We discuss "appropriate" distributions in Section 4.

and we save the residuals and the R-squared (which we denote by $R^2_{j,1}$) of this regression. Of course, if the sample size is large, it would be desirable to add more than one non-linear direction into this regression. We then run a regression of these residuals on the same regressors in addition to y_{t-p}, and $\phi\left(\delta'_j(1, y_{t-p})\right)$ where δ'_j is also randomly chosen. We save the R^2 of this second regression which we denote by $R^2_{j,2}$. The first measure of nonlinear partial autocorrelation that we propose is

$$nlpac_1\,(p) = \text{weighted average of } R^2_{j,2} \tag{8}$$

where the weights are proportional to $R^2_{j,1}$. If (5) was true and if the random one node neural network model was an adequate model of $E\,(y_t \mid y_{t-1}, ..., y_{t-p+1})$, then $TR^2_{j,2}$ would be an LM test statistic of the null hypothesis that, once we have the regressor set given by $\left\{1, y_{t-1}, ..., y_{t-p+1}, \phi\left(\gamma'_j(1, y_{t-1}, ..., y_{t-p+1})\right)\right\}$, then the additional regressors y_{t-p}, $\phi\left(\delta'_j(1, y_{t-p})\right)$ are not significant in explaining y_t. This LM statistic would have an asymptotic χ^2_2 distribution. However, even if (5) is true, some of the one node neural network models might carry only a faint signal about the conditional expectation of y_t given $y_{t-1}, ..., y_{t-p+1}$, and the test of the null that in those models $\{y_{t-p}, \phi\left(\delta'_j(1, y_{t-p})\right)\}$ adds no additional explanatory power might reject (5). This justifies weighting the second stage $R^2_{j,2}$ by the first stage $R^2_{j,1}$. Our simulations show that the critical values for this measure are similar to those based on $\frac{1}{T}\chi^2_2$.

The second measure of nonlinear partial autocorrelation that we propose is based on comparison of estimates of $E\,(y_t \mid y_{t-1}, ..., y_{t-p+1})$ and $E\,(y_t \mid y_{t-1}, ..., y_{t-p+1}, y_{t-p})$. We use a weighted average of projections onto random nonlinear directions as our estimate of these expectation. Specifically, we first compute

$$\hat{y}_{t|t-1,...,t-p+1} = \text{weighted average of projections of } y_t \text{ onto } \mathcal{S}_j(y_{t-1}, ..., y_{t-p+1})$$

where $\mathcal{S}_j(y_{t-1}, ..., y_{t-p+1})$ is the span of $\left\{1, y_{t-1}, ..., y_{t-p+1}, \phi\left(\gamma'_j(1, y_{t-1}, ..., y_{t-p+1})\right)\right\}$ and the weights are proportional to the variance of each projection. Again, if the sample size is large, it would be desirable to add more than one ϕ in the construction of $\mathcal{S}_j$ to minimize the effect of the approximation error. We compute $\hat{y}_{t|t-1,...,t-p+1,t-p}$ analogously, and measure the importance of additional lag p by looking at

$$nlpac_2\,(p) = r^2\left(y_t, \hat{y}_{t|t-1,...,t-p+1,t-p}\right) - r^2\left(y_t, \hat{y}_{t|t-1,...,t-p+1}\right). \tag{9}$$

This measure compares two approximate predictions for y_t based on the same procedure but with different number of lags. This measure is almost[9] equivalent to a comparison

[9] This equivalence is not exact because the average of orthogonal projections may not be a projection.

of the sum of squared prediction errors when using $t-p+1$ lags versus $t-p$ lags in predicting y_t, and hence it is similar to an unconditional test for forecast equivalence.

4 Empirical Performance of the Correlation Measures

Our simulation study is based on twelve data generating processes considered by Granger and Lin (1994) and Granger et al (2003). These processes are listed in Table 1, and they provide an interesting collection of processes to study because some are nonlinear MA processes, others are nonlinear AR processes, and many of them have lag structures that linear correlograms and partial correlograms are unable to detect. All simulations were performed using Gauss, and we based our study on samples of 100 and 200 observations.

Each element of γ_j and δ_j was drawn independently from appropriately chosen uniform distributions, with the total number of draws (i.e. m) set equal to 500. The range of the uniform distributions had to be chosen so that they were wide enough to give a reasonable coverage of the rectangle of height 1 over the range of y_{t-p}, but narrow enough to stay within the precision range of the computer's math coprocessor. Specifically, for

$$\phi\left(\gamma_j'(1,y_{t-p})\right) \equiv \frac{1}{1+\exp\left\{-a_j\left(y^{cn}_{t-p}-b_j\right)\right\}}$$

where y^{cn}_{t-p} is the centered and standardized y_{t-p}, then reasonable ranges for a_j and b_j are $[0,9]$ and $[-2,2]$.[10] Note that there is no need to consider negative values for a_j because

$$\frac{1}{1+\exp\left\{a_j\left(y^{cn}_{t-p}-b_j\right)\right\}} = 1-\frac{1}{1+\exp\left\{-a_j\left(y^{cn}_{t-p}-b_j\right)\right\}}.$$

For calculations relating to partial correlations, we needed to control the range of $\exp\left\{-[a_j{}'(y^{cn}_{t-1},\ldots,y^{cn}_{t-p})-a^*_j b_j]\right\}$, which we did by taking the first element of a from the $U[0,9]$ distribution, and the remainder from the $U[-9,9]$ distribution (so that the contributions from other lags were equally likely to increase or decrease the overall total of $a_j'y^{cn}$ relative to $a_{j1}y^{cn}_{t-1}$). Next, we drew b_j from the $U[-2,2]$ distribution and multiplied it by a^*_j, the standard deviation of $a_j{}'(y^{cn}_{t-1},\ldots,y^{cn}_{t-p})$. We then multiplied all terms in the exponent by a factor of $\frac{1}{\sqrt{p}}$, so that the variance of the sum of the p terms was approximately the same, regardless of the value of p.

[10] Since for a given value of b_j, large values of a_j produce functions that are very similar (i.e. highly correlated), one might want to give less probability to drawing larger values of a_j. For example, one could let $a_j = -9\ln\left(\mathcal{U}\left[e^{-1},1\right]\right)$, which amounts to drawing randomly from a truncated exponential distribution on $[0,9]$. Our simulations are based on $a_j = \mathcal{U}\,[0,9]$.

Full details of all simulated distributions for all correlation measures are available upon request, and we just report summary results for a few DGPs here. In order to give readers an idea of the distribution of each measure, we provide box and whisker plots, with the box outlining the 25th to 75th percentile ranges, and the whiskers stretching out from the 5th to 95th percentiles. For any given measure, each diagram contains two box and whisker plots for each lag, with the first plot showing the measured correlation for the DGP under consideration, and the second plots showing the same correlation measure for a white noise process.

4.1 Nonlinear Autocorrelograms

Figure 1 shows the performance of our second nonlinear autocorrelation measure (from equation 3), that is based on an unweighted average of the R^2s between y_t and predictions formed from regressing y_t on a constant, y_{t-p} and a squashed function of y_{t-p}. We provide four diagrams that relate to Models 2, 4, 6 and 9. The dominant feature of the top left hand diagram is the distribution relating to lag 2. This distribution is clearly centred well above zero, and well above the corresponding distribution for white noise. None of the other distributions on this diagram are different from distributions derived from white noise, so it is quite clear that the mean of this process has a dependency at lag 2, and no other lag dependencies.

Turning to the top right hand diagram that corresponds to Model 4, we see very good power in detecting nonlinear correlation up to lag 2, with lag 3 detected 25% of the time. The third diagram corresponds to Model 6, which is a nonlinear AR(1) process. Here $nlac_1(p)$ detects significant correlation for the first three or four lags and also shows a pattern of decay that is typical of autoregressive processes. The final diagram relates to a bilinear DGP, and here we see that the distributions for the first two correlation measures are a little higher and more variable that those for white noise, but $nlac_1(p)$ has little power in distinguishing this process from a white noise. Non-parametric measures of dependence, e.g. Granger et al (2003), also have difficulty in detecting dependence in this model.

The results for Model 2 are particularly encouraging, especially since standard correlograms do not indicate any lag structure for this DGP. Although we have not provided the relevant diagrams, our nonlinear correlograms also show lag structure (at lag 1, and then at lag 3) when applied to Models 1 and 3 respectively, and very pronounced lag 1 structure for Model 10 (see Figure 5), even though standard autocorrelation analysis fails to find any structure in any of these cases. The nonlinear autocorrelations for model

5 are very similar to those illustrated for Model 6, and those for Models 7 and 8 look very similar to their linear counterparts. Finally, our nonlinear correlograms find no lag structure for the conditional mean of the GARCH(1,1) process (Model 11), leading to the correct conclusion that one can not use lagged innovations to predict the mean of a GARCH(1,1) process.

The effect of excluding y_{t-p} in the prediction equation (i.e. using $nlac_1^0$ rather than $nlac_1$) is to lower the mean and standard deviation of all of the distributions of correlation measures. These changes are often quite substantial, (i.e. typically the removal of y_{t-p} from the regressor set reduces the measured correlation by about a half), but since the corresponding benchmark distributions based on white noise also experience the same sorts of change (as we move from a χ_2^2 to a χ_1^2), there is little change in what the diagrams tell us. The effect of increasing the sample size from 100 to 200 tightens the distributions, as expected.

Figure 2 illustrates the performance of our third correlation measure (i.e. $nlac_2$), in which we weight predictions by their R^2s, and then measure the correlation between y_t and its weighted predictions. The patterns in Figure 2 mimic those in Figure 1, although now the elevated distributions (e.g. the second lag measure for Model 2) differ more clearly from their counterparts based on white noise. Note that the scales on Figure 2 are bigger than those on Figure 1. As for the $nlac_1$ measures considered above, the omission of y_{t-p} in the prediction equation (i.e. using $nlac_2^0$ rather than $nlac_2$) lowers the mean and standard deviation of all of the distributions of correlation measures, but in this case, the changes are not substantial. As above, the effect of increasing the sample size from 100 to 200 is as expected.

We also experimented with two other measures $nlac_3^0$ and $nlac_3$, which were based on finding the correlation between y_t and a composite $\hat{y}_t$ formed by taking a simple average of all neural network functions of y_{t-p}. These results were all in between the results illustrated in Figures 1 and 2. Table 2 provides our estimated 5% critical values for all six correlation measures, for samples of size 100 and 200.

4.2 Partial Autocorrelograms

Figures 3 and 4 illustrate the distribution of our two partial correlation measures, for the same DGPs as in Figures 1 and 2. Both measures exhibit similar behaviour in measuring the partial correlations of the nonlinear MA processes, and these resemble the behaviour of correlations. This may seem surprising, but there is no reason to expect a slowly decaying partial autocorrelation in quadratic moving average processes such as those studied here. We focus our discussion in the rest of this section on nonlinear autoregressive processes, for which partial autocorrelation measures are likely to be most useful.

Figure 3 illustrates $nlpac_1$, the first of our partial autocorrelation measures, in which the R^2 from our second stage LM regression is weighted by the R^2 from the first stage. Interesting results from the diagrams are that for the nonlinear AR(1) process (Model 6), we see that the lag 1 distribution for our measure is well above its white noise counterpart, while none of the measures for any of the other lags show any evidence of statistical significance. Thus the partial correlogram has correctly identified that we need only the first lag of y_t (i.e. y_{t-1}) to predict this process. We have not illustrated the nonlinear partial correlograms for the other AR(1) processes (i.e. Models 5, 7, and 8), but these partial correlograms exhibit exactly the same behaviour. Our $nlpac_1$ measure shows no significant partial correlation for the bilinear process (see the fourth panel in Figure 3). Finally, as we might hope, our nonlinear partial correlogram finds no lag structure for the conditional mean of the GARCH(1,1) process (Model 11).

Critical values for $nlpac_1$ coefficients increase very slightly with lag length, as the first stage R^2 becomes more variable. However, these changes are likely to be negligible in practical situations, and simulations based on 5000 DGPs with $m = 500$ lead to 5% critical values of 0.054 for samples of 100 and 0.027 for samples of 200.

We illustrate our other partial correlation measure ($nlpac_2$) in Figure 4. This measure is based on the difference between the R^2 of the nonlinear prediction of y_t based on $y_{t-1}, ..., y_{t-p}$ and the R^2 of the nonlinear prediction of y_t based on $y_{t-1}, ..., y_{t-p+1}$. Unlike $nlpac_1$ this partial autocorrelation measure can be negative, if the addition of the lag p variables causes a deterioration in forecastability (see the conclusion for a discussion). Also, critical values decline very rapidly with lag length, which complicates the use of this measure and makes it impractical once long lag lengths are considered. However, this measures performs well for the DGPs studied here. In particular, as is evident in the fourth panel of Figure 4, this measure identifies predictability at lag 2 for the bilinear model. This second lag effect is clearly statistically significant, and shows that $nlpac_2$ is able to correctly identify subtle nonlinear structures that traditional correlation tools cannot find.

Finally, in Figure 5 we present the performance of all four measures when applied to a sample of 100 observations from a chaotic process (Model 10, the "tent map"). This is a deterministic nonlinear process which cannot be distinguished from white noise by linear autocorrelogram and partial autocorrelogram. The two measures of nonlinear autocorrelation find significant correlation at lags 1 and 2, and in particular, the distribution of autocorrelation at the first lag is concentrated so close to 1 (it has a mean of 0.9968 and a standard deviation of 0.001) that its box and whisker plot is a point on the graph. While $nlpac_1$ still shows a significant partial correlation at lag 2, $nlpac_2$ correctly identifies this as an autoregressive process of order 1.

4.3 Applications

In order to determine the usefulness of our measures, we firstly check their performance with respect to some nonlinear DGPs that have been used in the applied econometrics literature to model unemployment and industrial production, and then we analyse the data that was actually used to estimate these models.

Nonlinear autoregressive models are very popular in the applied literature, and here we focus on Rothman's (1998) threshold (TAR) model of US unemployment, and the smooth transition autoregressive (STAR) models of industrial production for Belgium and Japan, taken from Teräsvirta and Anderson (1992). The lag lengths associated with these models are 2, 5 and 9 respectively. Full descriptions of the models are provided in Table 3, and Figures 6 to 8 show how well our nonlinear autocorrelograms and partial autocorrelograms would perform, given samples of size 100, generated from each of these three models.

These figures are very reassuring, since the distributions of the partial autocorrelation coefficients are clearly above the corresponding white noise distributions for up to two lags for the TAR(2) model, for up to five lags for the logistic L-STAR(5) model, and for up to nine lags for the exponential E-STAR(9) model. Thus, it appears that our nonlinear partial autocorrelation functions can pick the lag length in quite complicated nonlinear autoregressive processes, even if the lag length is quite long. The distributions of the nonlinear autocorrelation coefficients are well above their white noise analogues, even for quite large lag lengths, in line with what we would expect given the autoregressive structure of the data.

Table 4 provides the standard and nonlinear correlation and autocorrelation functions for the data that was actually used to estimate the models. The linear and nonlinear measures "agree" in the first two cases, selecting a two lag autoregressive process in the first and a five lag autoregressive process in the second. These results are supported by both AIC and BIC (applied to linear autoregressions) in the first case, and by BIC in the second (AIC chooses 8 lags in the second case). One might therefore conclude that for these DGPs the nonlinear measures provide no additional information to standard lag selection techniques.

The story changes, however, once we look at the Japanese Industrial Production data. AIC and BIC based on linear autoregressions each choose five lags, and this choice is supported by the standard correlation and autocorrelation functions. However, despite this strong support for five lags, Teräsvirta and Anderson (1992) found that they needed nine lags to build an appropriate nonlinear model of this data. It is interesting to note, then, that our $nlpac_1$ measure "finds" structure at lag 9 for the Japanese series, as was found necessary by Teräsvirta and Anderson (1992). The statistically significant measures for $nlac_2^*$ (at lags nine and ten) are also quite consistent with a deep lag structure for this series. While we have not reported the details, the story is the same (but not quite as clear) for the Italian index of industrial production studied by Teräsvirta and Anderson (1992). Standard techniques used for linear processes "found" five lags, yet nine lags were need to build an appropriate nonlinear model.

We do not know the true lag structure for these series since they were generated by nature, but we know that attempts to fit nonlinear models with five lags led to models that were clearly mispecified, while attempts to fit nonlinear models with nine lags produced models that passed specification tests. Also, we know from Teräsvirta and Anderson (1992) that the nine lag nonlinear models produced slightly better out-of-sample forecasts than the linear AR(5) models, which offers further support for the longer lag structure. Our examples demonstrate that our nonlinear autocorrelograms and partial autocorrelograms can dominate their linear analogues when the data contain nonlinearities, and this suggests that they might provide useful tools for specifying lag lengths in nonlinear time series models.

5 Conclusion

This paper studies the problem of lag selection in nonlinear models. We develop a neural network based method for calculating dependence in conditional mean, and then use this method to construct nonlinear analogues to autocorrelograms and partial autocorrelograms. While there are several nonlinear autocorrelograms that are currently available, ours are easier to calculate, and they seem to work well in relatively small samples. There are very few nonlinear analogues to the partial correlogram that are currently available[11], and we believe that ours are perhaps the first that are practical enough to be used by applied researchers. Given the importance of nonlinear AR processes in the

[11] One exception is Kendall's (1938) partial τ.

applied econometrics literature, and the fact that for AR processes the partial correlograms are more useful for identifying lag length than correlograms, we believe that our nonlinear partial autocorrelograms are likely to be useful. Interested researchers may contact us for our GAUSS programs.

Our Monte Carlo study is an exploratory investigation of whether the crude one node neural network based measures were of any value. Given the promising results of this study, there is much room for improving these measures. Possible directions for improvement are:

1. Increasing the number of nodes: We chose one node to see if the crudest neural network approximation has any power in detecting lag dependence. We have no reason to discourage users from using multiple squashing functions, especially when analyzing large samples. Since random draws of γ_j may lead to highly correlated ϕ_j, it is common practice in neural network based tests for nonlinearity (see Teräsvirta et al 1993) to form many (say 10) ϕ_j and choose the first few (say 3) principal components of them. We think that this may improve our measures further, especially since the first partial correlation measure is only theoretically justified if the neural network approximation of $E\left(y_t \mid y_{t-1}, ..., y_{t-p+1}\right)$ is an accurate one.

2. Increasing m (the number of predictive directions) as p (the number of lags) increases: We averaged over the same number of predictive directions as we increased p. This often led to a smaller R^2 as p increased, and made $nlpac_2$ negative. This may be disconcerting to some, as the model with p lags nests the model with $p-1$ lags. We don't see this apparent anomaly as a major problem, but we think that it can be avoided by increasing m with p.

We believe that one reason that our measures are successful is that they combine forecasts that each carry a faint signal about the nonlinear relationship between y_t and its lags. Interpreting this nonlinear relationship as a common factor that characterizes each set of predictors that is used to obtain our aggregate measure of predictability, then averages of our predictions provide estimates of this common factor and our aggregate measures of predictability incorporate this common factor. This interpretation falls in line with Granger and Jeon (2002), who link the superior ability of combined forecasts to the presence of common factors.

References

Akaike, H. (1974), "A New Look at Statistical Model Identification", *IEEE Transactions on Automatic Control, AC19*, 716-723

Anderson, H. (2002), "Choosing Lag Lengths in Nonlinear Dynamic Models", forthcoming in **Advances in Economics and Econometrics**, Edward Elgar Press.

Auestad, B. and D. Tjøstheim (1990), "The Identification of Nonlinear Time Series: First Order Characterization and Order Determination", *Biometrika* (1990), 77, 669 - 687.

Bierens, H. J. (1990), "A Consistent Conditional Moment Test of Functional Form", *Econometrica*, 58, p 1443 - 1458.

Brock, W., W. Deckert, J. Scheinkman and B. LeBaron (1996), "A Test for Independence Based on the Correlation Dimension", *Econometric Reviews* 15(3), 197-236.

Boivin J. and S. Ng, (2002), "Are More Data Always Better for Factor Analysis? ", working paper, John Hopkins University.

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

Elliott, G. and A. Timmermann (2002), "Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions", University of California at San Diego Working Paper 2002-08.

Gourieroux, C. and J. Jasiak (2002), "Nonlinear Autocorrelograms; An Application to Inter-Trade Durations", *Journal of Time Series Analysis*, 23, 127 - 154.

Granger, C.W.J. (1987), "Implications of Aggregation with Common Factors" *Econometric Theory*, 3, 208 - 222.

Granger, C.W.J. (1989), "Invited Review: Combining Forecasts - Twenty Years Later", *Journal of Forecasting*, 8, 167 - 173.

Granger, C.W.J. and Y. Jeon (2002), "Combined Forecasts and Hidden Common Factors", mimeo, University of California at San Diego.

Granger, C.W.J. and J. Lin (1994). "Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models", Journal of Time Series Analysis, 15, 371 - 384.

Granger, C.W.J., E. Maasoumi and J. Racine, (2003), "A Dependence Metric for Possibly Nonlinear Processes", mimeo, University of Syracuse, New York.

Hannan, E.J., and B. G. Quinn (1979), "The Determination of the Order of an Autoregression,", Journal of the Royal Statistical Society, Series B, 41, 190 - 195.

Hansen, B. (1992), "The Likelihood Ratio Test under Non-standard Conditions: Testing the Markov Switching Model of GNP", Journal of Applied Econometrics, 7, 661-682.

Hendry, D.F. and M.P. Clements (2002), "Pooling of Forecasts", Econometrics Journal, 5, 1 - 26.

Hong, Y. and T. Lee (2003), "Diagnostic Checking for the Adequacy of Nonlinear Time Series Models", forthcoming in Econometric Theory.

Hong, Y. and H. White (1995), "Consistent Specification Testing via Non-parametric Series Regression, Econometrica, 63, 1133-1160.

Hong, Y. and H. White (2000), "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence", UCSD, working paper available at http://www.econ.ucsd.edu/faculty/

Hornik, K.M., M. Stinchcombe and H.L. White, (1989). "Universal Approximations of an Unknown Mapping and its Derivatives using Multi-Layer Feedforward Networks", Neural Networks 3, 551 - 560.

Keenan, D. M. (1985), "A Tukey Non-Additivity Type Test for Time Series Nonlinearity", Biometrika, 72, 39 - 44.

Kendall, M. (1938), "A New Measure of Rank Correlation ", Biometrika, 30, 81-89.

Lee, T-H, H. White and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests", Journal of Econometrics, 56, 268-290.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", Journal of the Royal Statistical Society, B, 31, 350 - 371.

Rothman, P. (1998), “Forecasting Asymmetric Unemployment Rates”, Review of Economics and Statistics, 80, 164 - 168.

Schwartz, G. (1978), “Estimating the Dimension of a Model”, The Annals of Statistics, 6, 461 - 464.

Stock J.H. and M. W. Watson (2000), “Forecasting Output and Inflation: the Role of Asset Prices”, Working paper, Princeton.

Teräsvirta, T. and H.M. Anderson (1992), “Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models”, Journal of Applied Econometrics, 7, S119-S136.

Teräsvirta, T. (1994), “Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models”, Journal of the American Statistical Association, 89, 208-218.

Teräsvirta, T., C.F. Lin and C. W. J. Granger (1993), “Power of the Neural Network Linearity Test”, Journal of Time Series Analysis, 14, 209 - 220.

Teräsvirta, T., D. Tjøstheim and C. W. J. Granger (1994), “Aspects of Modelling Nonlinear Time Series”, Chapter 48, **Handbook of Econometrics**, Volume 4, Editors, R.F. Engle and D.L. McFadden, North Holland.

Tjøstheim, D. and B. Auestad (1994). “Nonparametric Identification of Nonlinear Time Series: Selecting Significant Lags”, Journal of the American Statistical Association, 89, 1410 - 1419.

Tsay, R. (1989) “Testing and Modeling Threshold Autoregressive Processes”, Journal of the American Statistical Association”, 84, 231-240.

Table 1: DGPs used in the simulation studies

Model 0 : $y_t = \varepsilon_t$

Model 1 : $y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

Model 2 : $y_t = \varepsilon_t + 0.8\varepsilon_{t-2}^2$

Model 3 : $y_t = \varepsilon_t + 0.8\varepsilon_{t-3}^2$

Model 4 : $y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2 + 0.8\varepsilon_{t-2}^2 + 0.8\varepsilon_{t-3}^2$

Model 5 : $y_t = |y_{t-1}|^{0.8} + \varepsilon_t$

Model 6 : $y_t = sign(y_{t-1}) + \varepsilon_t$

Model 7 : $y_t = 0.8y_{t-1} + \varepsilon_t$

Model 8 : $y_t = y_{t-1} + \varepsilon_t$

Model 9 : $y_t = 0.6\varepsilon_{t-1}y_{t-2}+ \varepsilon_t$

Model 10: $y_t = 4y_{t-1}(1 - y_{t-1})$

Model 11: $y_t = \sqrt{h_t}\varepsilon_t$, $h_t = 0.01 + 0.94h_{t-1} + 0.05y_{t-1}^2$

(In all models $\varepsilon_t \sim N(0,1)$).

Table 2: 5% Critical values for nonlinear correlation coefficients

Sample	Measure					
Size	$nlac_1^0$	$nlac_2^0$	$nlac_3^0$	$nlac_1$	$nlac_2$	$nlac_3$
100	0.028	0.064	0.059	0.052	0.080	0.072
200	0.014	0.034	0.031	0.026	0.041	0.036

Critical values are based on 5000 DGPs, with $m = 500$.

Table 3: Nonlinear autoregressive models

TAR(2) $y_t = 0.0529 + 1.349 y_{t-1} - 1.665 y_{t-2} + f_t \times (1.646 y_{t-1} - 0.733 y_{t-2}) + \varepsilon_t$ with
$f_t = (1)(y_{t-1} < 0.062)$ and $\varepsilon_t \sim N(0, 0.063^2)$

LS(5) $y_t = -0.030 + 0.64 y_{t-1} - 0.29 y_{t-2} - 0.64 y_{t-4} +$

$f_t \times (0.044 + 0.49 y_{t-2} + 0.45 y_{t-5}) + \varepsilon_t$
with $f_t = (1 + \exp\{-7.3 \times 21.6 (y_{t-1} + 0.015)\})^{-1}$ and $\varepsilon_t \sim N(0, 0.0231^2)$.

ES(9) $y_t = 0.0075 + 3.03 y_{t-1} - 1.31 y_{t-2} - \Delta 0.49 y_{t-4} +$

$$f_t \times (-1.68 y_{t-1} + 0.87 y_{t-2} - \Delta 0.30 y_{t-8}) + \varepsilon_t \text{ with}$$
$$f_t = (1 - \exp\{-1.54 \times 196 (y_{t-1} + 0.082)^2\}) \text{ and } \varepsilon_t \sim N(0, 0.0185^2).$$

Table 4: Performance of linear and nonlinear measures of autocorrelation and partial autocorrelation

Detrended employment for the United States (sample of 128)

Correlation	Critical	Lag Length									
Measure	Value (5%)	1	2	3	4	5	6	7	8	9	10
ac	±.18	.94*	.79*	.62*	.45*	.31*	.22*	.15	.10	.06	.03
pac	±.18	.94*	-.67*	.09	.06	.13	-.09	-.10	.04	.09	-.19*
$nlac_1$	.05	.88*	.65*	.42*	.25*	.14*	.08*	.05	.02	.01	.00
$nlac_2$	.08	.88*	.65*	.43*	.27*	.17*	.11*	.07*	.05*	.05	.05
$nlpac_1$	.05	.88*	.43*	.02	.01	.02	.01	.01	.01	.01	.04
$nlpac_2$	Varies	.88*	.05*	.00	.00	.00	.00	.00	.00	.00	.00

Annual Growth Rate of Belgian Industrial Production (sample of 104)

Correlation	Critical	Lag Length									
Measure	Value (5%)	1	2	3	4	5	6	7	8	9	10
ac	±.21	.75*	.49*	.18	-.12	-.15	-.16	-.06	.02	.02	.05
pac	±.21	.75*	-.15	-.30*	-.25*	.42*	-.10	.03	-.17	.09	.02
$nlac_1$	.05	.59*	.25*	.04	.03	.04	.06*	.03	.03	.04	.02
$nlac_2$	.08	.59*	.25*	.06	.04	.04	.07	.07	.08	.07	.04
$nlpac_1$	.05	.59*	.10*	.13*	.07*	.18*	.05	.01	.04	.03	.01
$nlpac_2$	Varies	.59*	.03	.04	.02	.06*	.00	.00	.00	.00	.00

Annual Growth Rate of Japanese Industrial Production (sample of 104)

Correlation	Critical	Lag Length									
Measure	Value (5%)	1	2	3	4	5	6	7	8	9	10
ac	±.21	.91*	.71*	.45*	.19	.00	-.11	-.13	-.07	.05	.17
pac	±.21	.91*	-.63*	-.16	-.06	.28*	.15	-.04	.03	.19	-.07
$nlac_1$	.05	.83*	.51*	.21*	.05*	.01	.02	.02	.02	.02	.04
$nlac_2$	.08	.83*	.51*	.22*	.06	.04	.04	.04	.06	.10*	.16*
$nlpac_1$	.05	.83*	.47*	.04	.01	.14*	.02	.02	.02	.08*	.02
$nlpac_2$	Varies	.83*	.08*	.00	.00	.00	.00	.00	.00	.00	.00

Figure 1: The measure of non-linear autocorrelation $nlac_1$ for selected DGPs

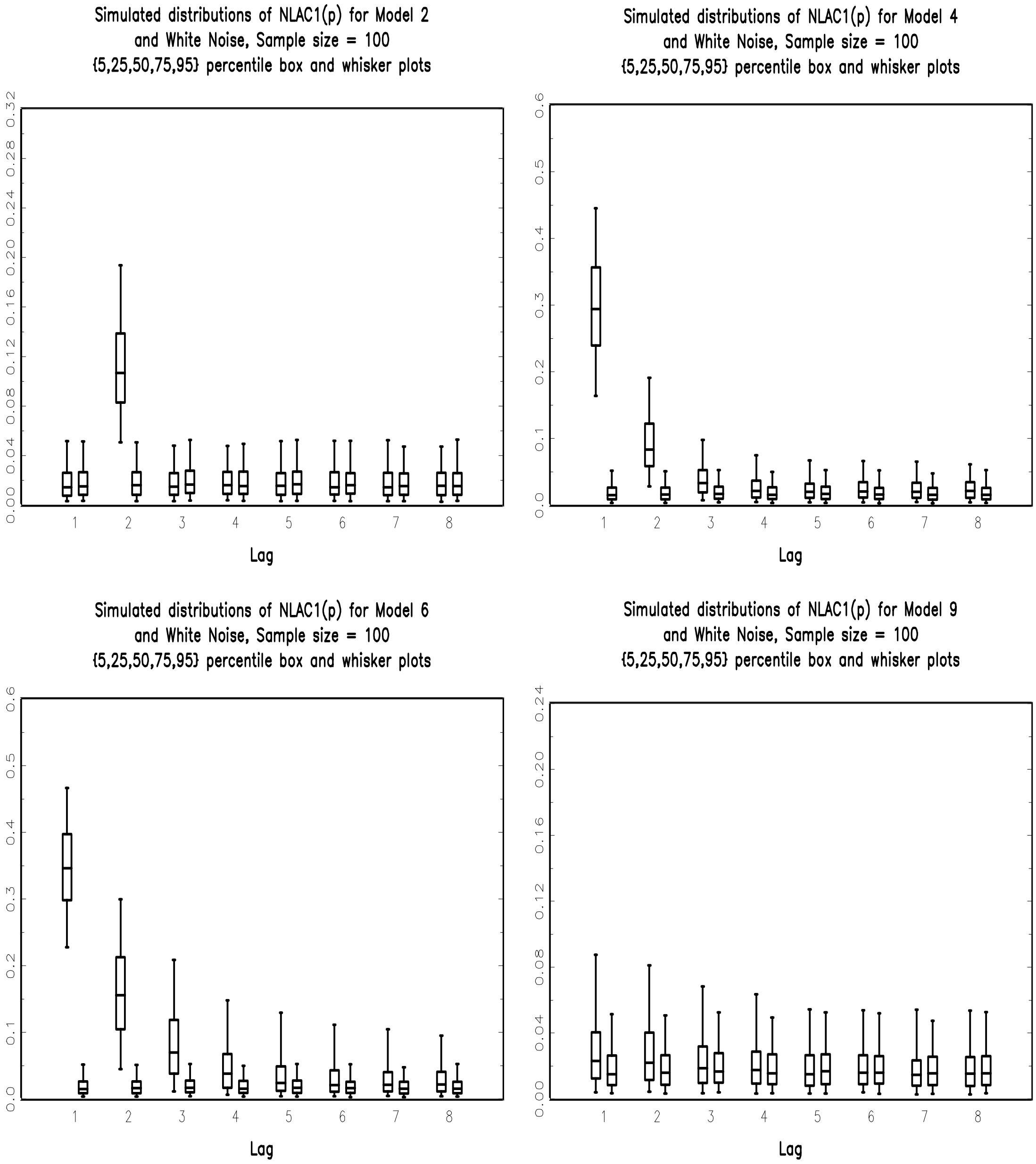
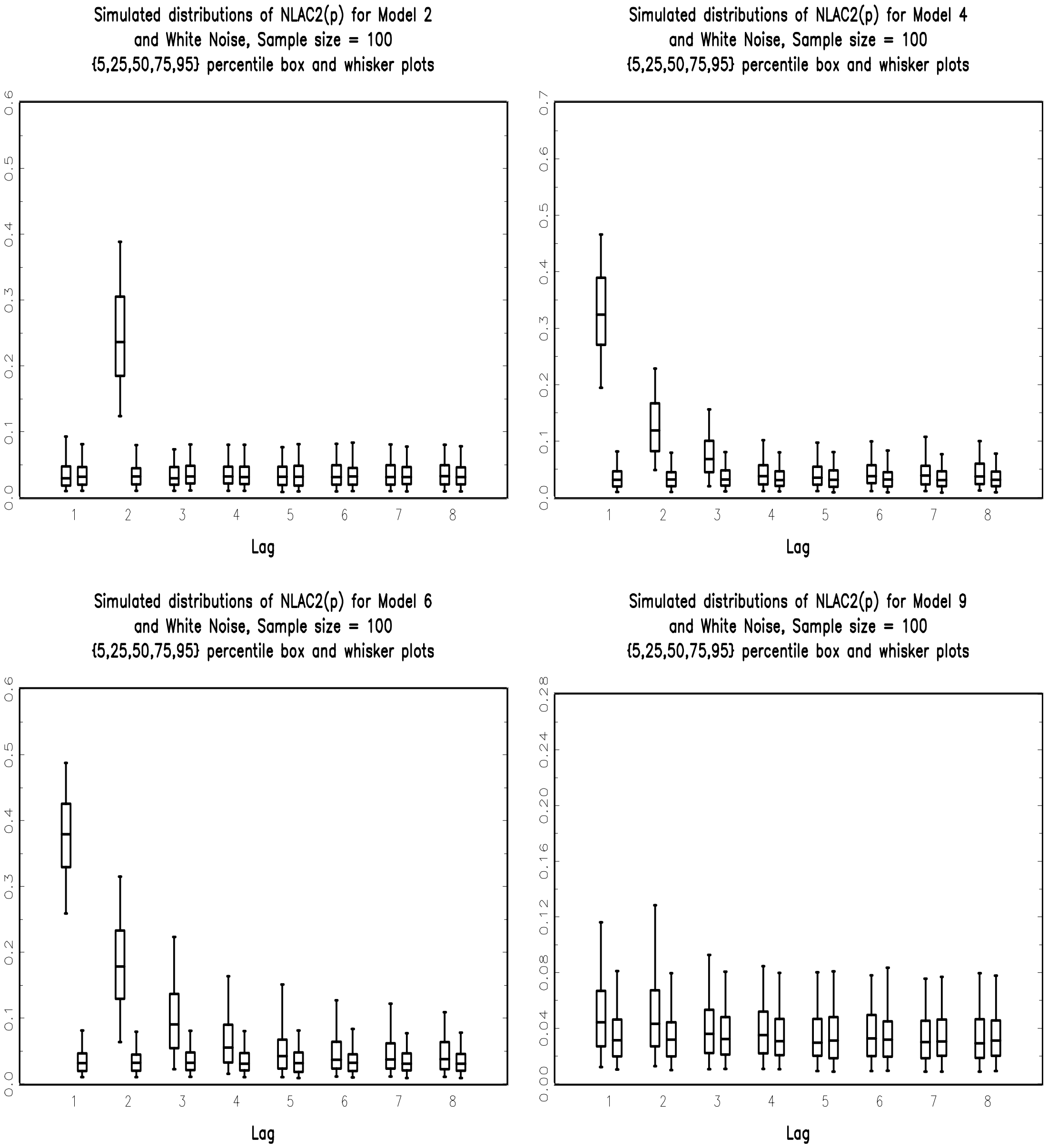


Figure 2: The measure of non-linear autocorrelation $nlac_2$ for selected DGPs


Simulated distributions of NLAC2(p) for Model 2 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) for Model 4 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) for Model 6 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) for Model 9 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Figure 3: The measure of non-linear partial autocorrelation $nlpac_1$ for selected DGPs

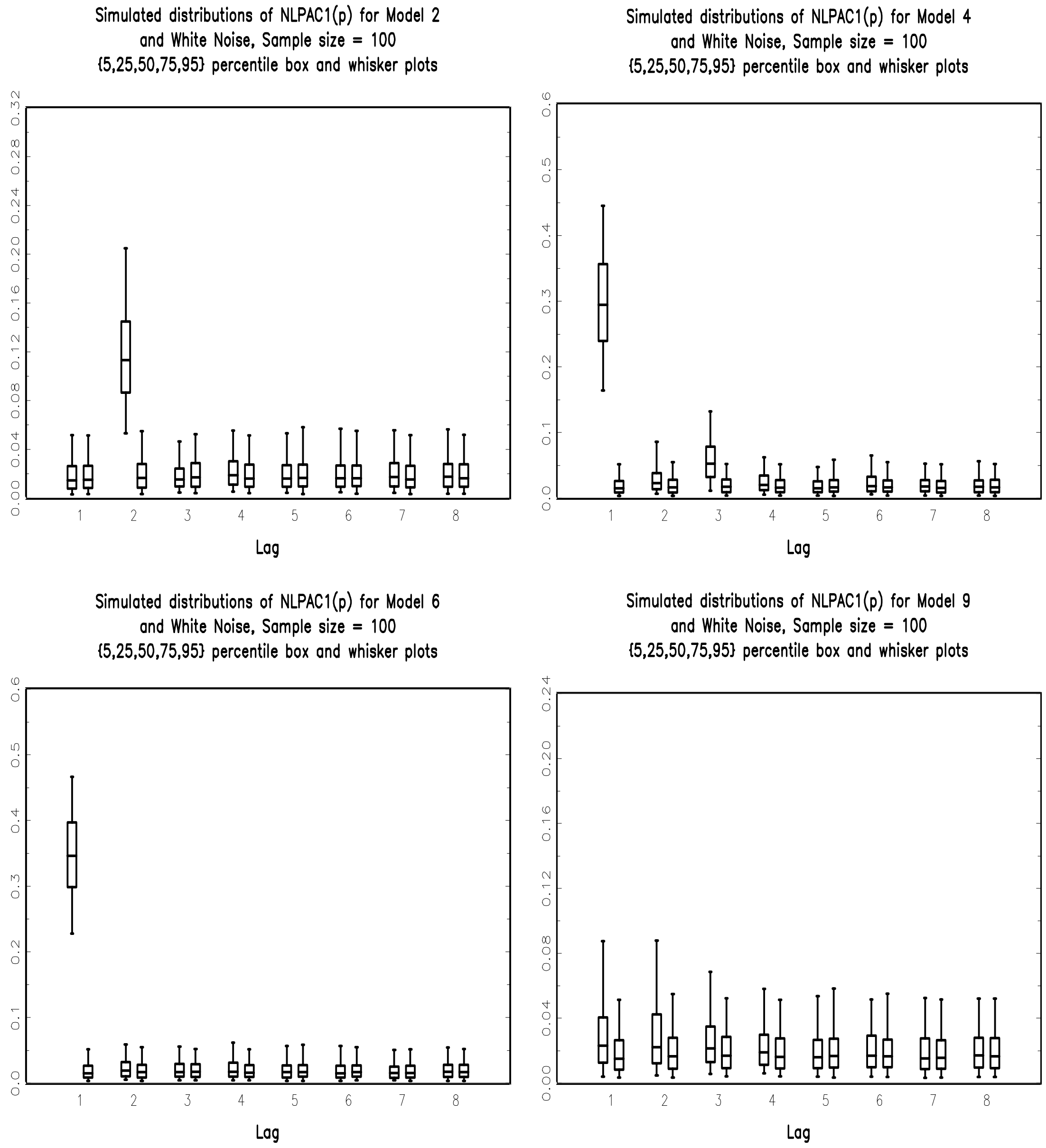


Figure 4: The measure of non-linear partial autocorrelation $nlpac_2$ for selected DGPs

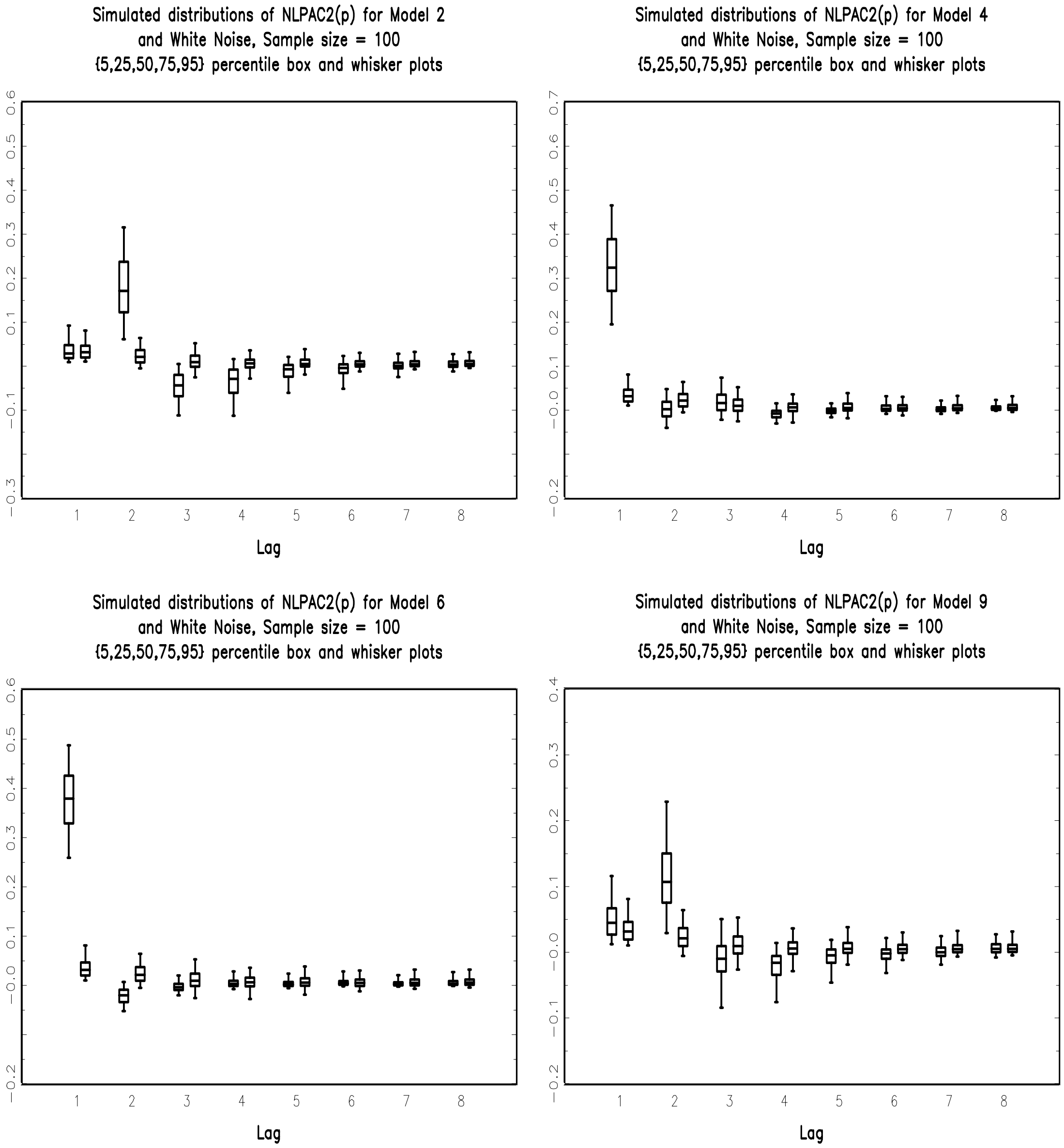
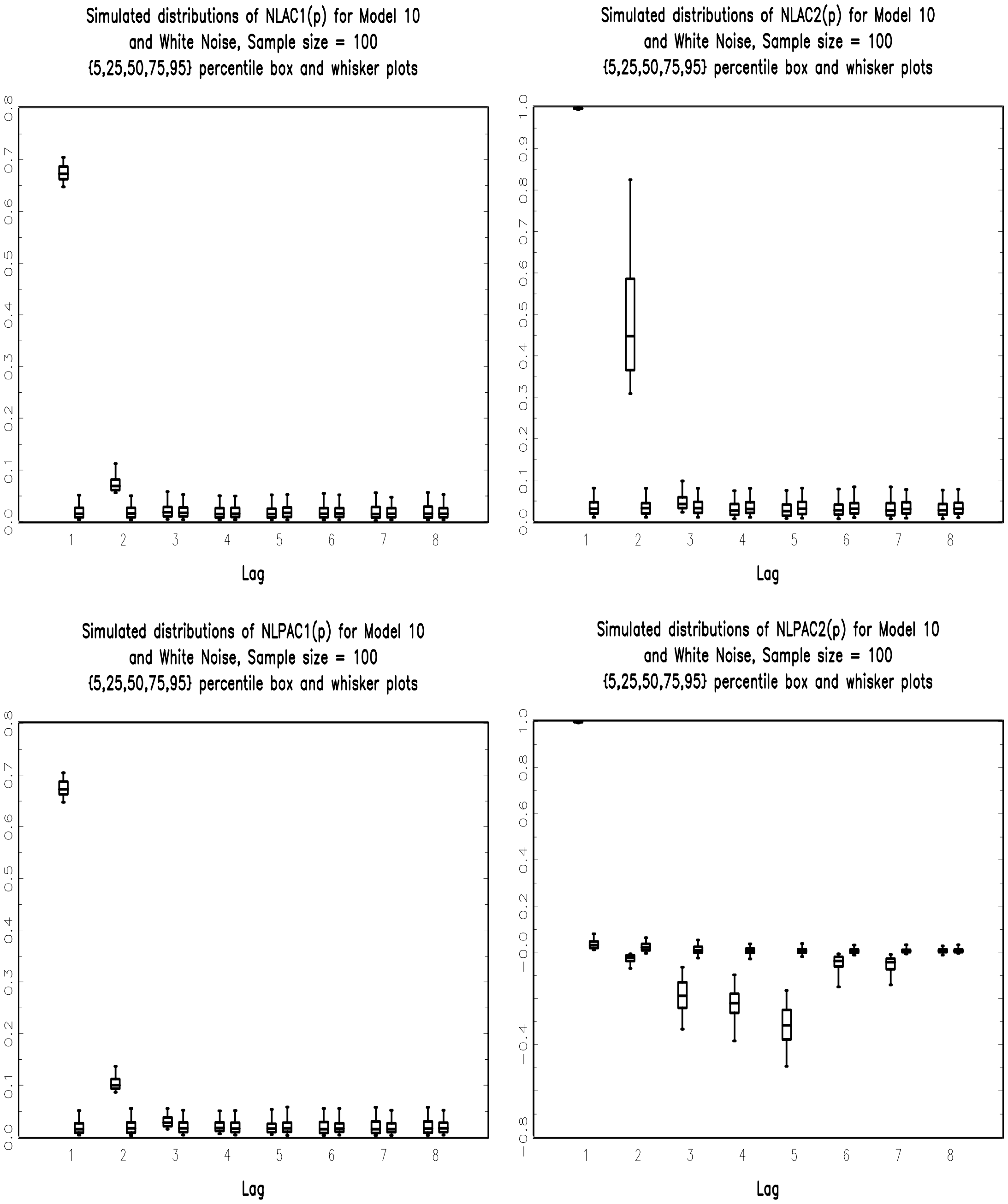


Figure 5: The measure of non-linear correlation and partial autocorrelation for the tent map



Simulated distributions of NLAC1(p) for Model 10 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) for Model 10 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC1(p) for Model 10 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC2(p) for Model 10 and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Figure 6: The measures of non-linear correlation and partial autocorrelation for the TAR(2) DGP

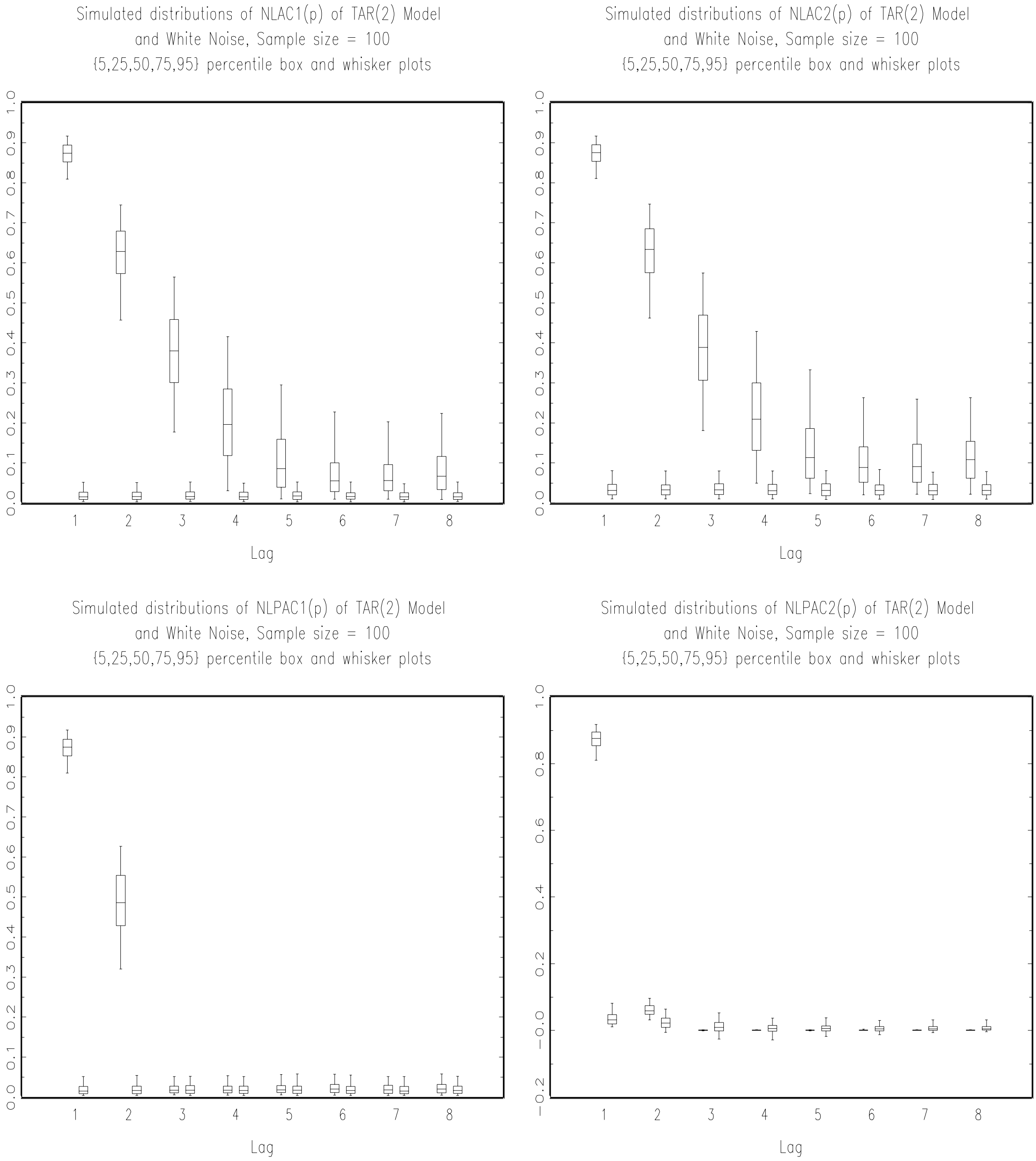
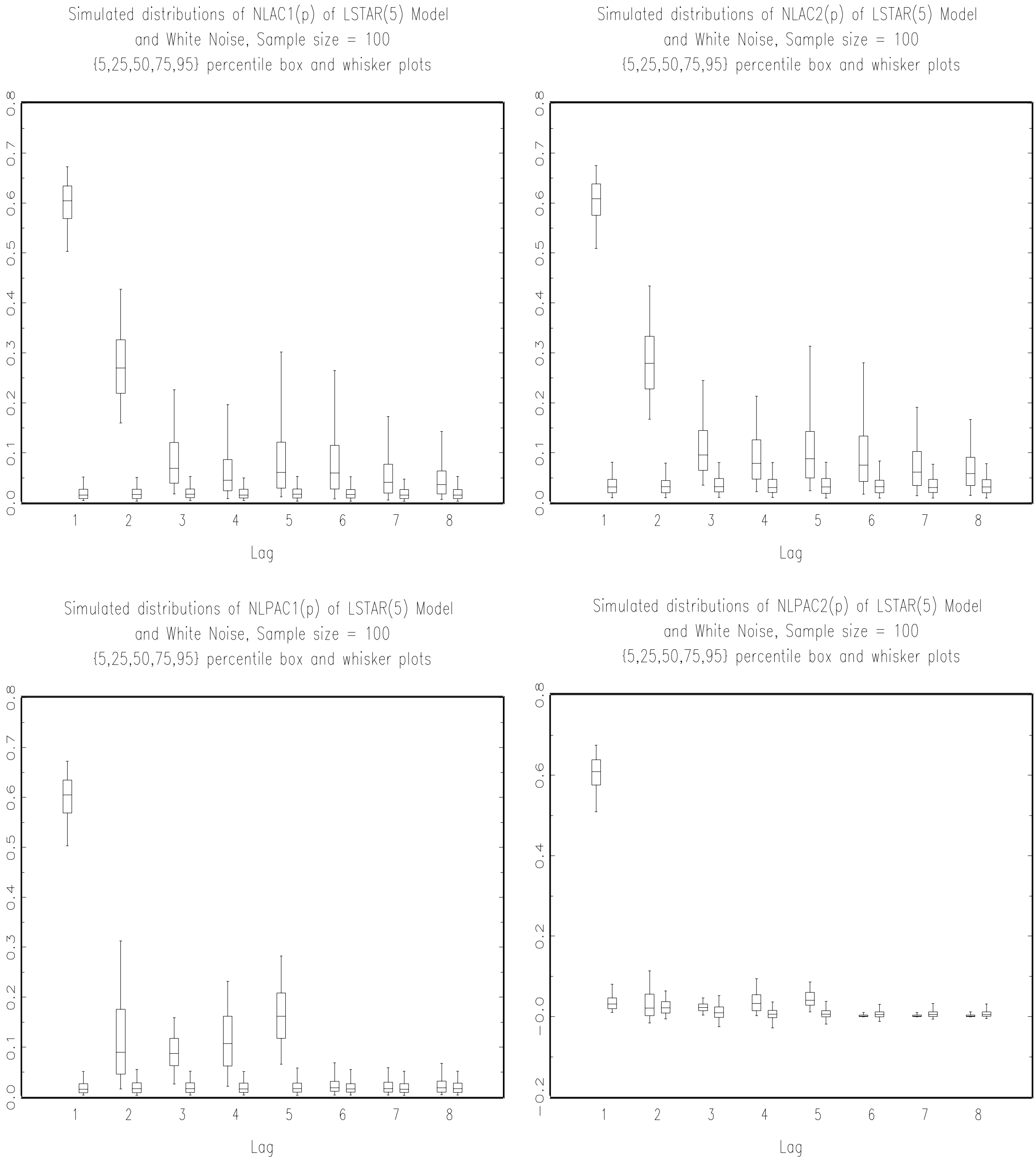


Figure 7: The measures of non-linear correlation and partial autocorrelation for the LSTAR(5) DGP



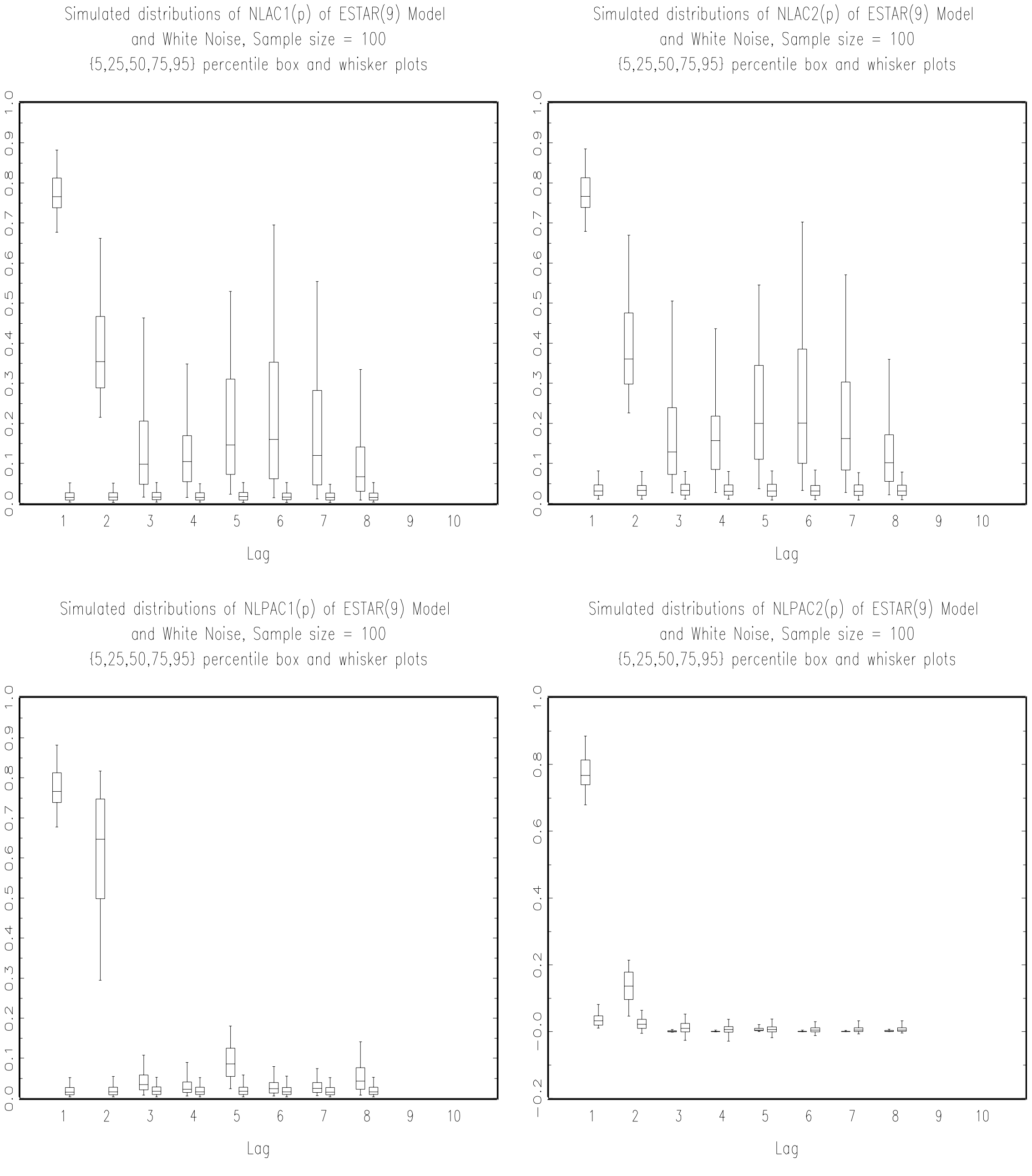
Simulated distributions of NLAC1(p) of LSTAR(5) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) of LSTAR(5) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC1(p) of LSTAR(5) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC2(p) of LSTAR(5) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Figure 8: The measures of non-linear correlation and partial autocorrelation for the ESTAR(9) DGP



Simulated distributions of NLAC1(p) of ESTAR(9) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLAC2(p) of ESTAR(9) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC1(p) of ESTAR(9) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots

Simulated distributions of NLPAC2(p) of ESTAR(9) Model and White Noise, Sample size = 100 {5,25,50,75,95} percentile box and whisker plots