Improved Accuracy of Asymptotic Autoregressive Theory

P.M.T. Broersen and H.E. Wensink

Signal Processing Group, Department of Applied Physics, Delft University of Technology
P.O. Box 5046, 2600 GA Delft, The Netherlands
e-mail: broersen@duttncb.tn.tudelft.nl
e-mail: wensink@duttncb.tn.tudelft.nl

RÉSUMÉ

La théorie asymptotique, qui est valable pour de grand nombre d'observations, décrit les résultats pratiques des estimateurs d'une précision d'ordre *1/N*. Pour déduire l'espérance statistique d'un estimateur il suffit de l'exprimer comme une fonction des covariances mesurées et ensuite de faire un développement de Taylor de cette fonction. La précision de cette théorie n'est pas améliorée en incorporant des termes d'ordre supérieurs, comme $1/N^2$ etc. Pour des descriptions plus précises, il faut utiliser les concepts de «Finite Sample Theory».

ABSTRACT

In autoregressive inference the asymptotic Large Sample Theory is currently used to describe the behaviour of the sample statistics. In investigating the limiting behaviour of sample statistics, probability concepts and Taylor approximations are important tools. To describe the statistical expectations with the accuracy of order *1/N*, a Taylor expansion of the quantity, written as a function of estimated covariances, is all that is required. Improvements to this accuracy of order *1/N* are not found by including higher order terms like $1/N^2$, but require the totally different concept provided by the Finite Sample Theory.

INTRODUCTION

Recently we have presented [1,2] a survey of three ways in which the estimation of an autoregressive (AR) model can theoretically be approximated (see Survey below). The essential difference between these levels is the way the residual variance, $S^2(p)$, and the squared error of prediction, $PE(p)$, are described. The $PE(p)$ indicates the forecasting capacities of the model, while $S^2(p)$ is the fit to the available data that have been used to estimate the model parameters. In AR order selection it is important to make a clear distinction between these quantities, since the accuracy of their descriptions may have a major influence on the selection results.

At first level the probability limits (Plim) [3] describe the classical results of AR theory. In the Plim the joint limiting distribution of the parameters equals the distribution of classical linear regression theory. As a result AR estimation is treated as a classical linear regression problem, when the sample size N is sufficiently large. The estimated parameters equal their expectation, because the bias disappears for $N \to \infty$ and is therefore not described at this level of approximation. Furthermore $S^2(p)$ equals the variance of the innovations ε_n that originally generated the process. There is no difference between predictions and residuals, since asymptotically their probability density functions are equal.

The second, asymptotical, level of approximation provides more detailed theoretical descriptions of the estimates that are accurate up to magnitude-order *1/N* (see S4-the S. numbers refer to formulae in the survey-). This degree of detail is minimally required to distinguish the different behaviour of $PE(p)$ [4,5] and $S^2(p)$. Furthermore it is possible to describe the bias in the estimates. The approximations at this level are found by considering parameters, residual variances and prediction errors as non-linear transformations of the measured covariances. A Taylor expansion of the resulting functions gives the desired descriptions with an accuracy that includes all terms of order *1/N* in the expectations of the stochastic functions. This paper demonstrates that we are not aware of a mathematical sound derivation of Taylor results with regression of the first theoretical level as starting point. Only unmotivated approximations can lead to some well-known formulae.

At the third level the way the estimation methods affect the estimation results is taken into account. Simulations show the important differences between the outcomes of estimation methods. The first two levels of approximation to AR theory contain no possibilities to adapt to the practical behaviour in finite samples. Four methods [6] have been evaluated: the Yule-Walker method (YW), the method of Burg (Burg), the Least Squares method that minimizes both Forward and Backward residuals (LSFB) and the Least Squares method that minimizes forward residuals only (LSF). Besides correcting for the estimation method, multiplicative formulae are used at this level instead of the additive asymptotical ones.

AR ESTIMATION

An autoregressive (AR) process of order K is defined as follows:

$$x_n + \sum_{i=1}^{K} a_i(K)\, x_{n-i} = \varepsilon_n, \qquad (1)$$

where ε_n is the generating innovation process, i.i.d. with zero mean, variance σ_ε^2 and finite fourth order moments. The theoretical parameters are the entries $a_i(K)$ of the parameter vector $\boldsymbol{a}(K)$. The word process denotes the true underlying process of order K, that need not be finite. To the data generated by the AR(K) process an AR model of order p can be fitted:

$$x_n + \sum_{i=1}^{p} \hat{a}_i(p)\, x_{n-i} = \hat{\varepsilon}_n, \qquad (2)$$

Survey: Three Levels of Approximation to Theory for AR(p) Model Estimation

Probability Limits	Large Sample Theory	Finite Sample Theory
ML theory	asymptotical justifications	practical data
$\lim_{N\to\infty} \mathcal{P}(\lvert\hat{x}_N - x\rvert \le \delta) = 1$, for any $\delta > 0$ (S1)	$\lim_{N\to\infty} N\, o(1/N) = 0$ (S4)	$v(i,YW) = (N-i)/N(N+2)$ $v(i,Burg) = 1/(N+1-i)$ $v(i,LSFB) = 1/(N+1.5-1.5i)$ $v(i,LSF) = 1/(N+2-2i)$ (S9)
$\underset{N\to\infty}{plim}\, S^2(p) = \sigma_\varepsilon^2$ (S2)	$\mathcal{E}[S^2(p)] = \sigma_\varepsilon^2(1 - p/N)$ (S5)	$\mathcal{E}[S^2(p)] = \sigma_\varepsilon^2 \prod_{i=0}^{p} [1 - v(i,.)]$ (S10)
$p \ge K$	$p \ge K$	$p \ge K$
$\underset{N\to\infty}{plim}\, PE(p) = \sigma_\varepsilon^2$ (S3)	$\mathcal{E}[PE(p)] = \sigma_\varepsilon^2(1 + p/N)$ (S6)	$\mathcal{E}[PE(p)] = \sigma_\varepsilon^2 \prod_{i=0}^{p} [1 + v(i,.)]$ (S11)
no order selection	$FPE(p) = S^2(p)\frac{N+p}{N-p}$ (S7)	$FSC(p) = S^2(p)\prod_{i=0}^{p}\frac{1+v(i,.)}{1-v(i,.)}$ (S12)
.....	.....	.....
	$GIC(p,\alpha) = \ln[S^2(p)] + \alpha p/N$ $\alpha = 2$ *AIC(p)* $\alpha = \ln(N)$ *consistent* $\alpha = ln\ln(N)$ *minimal consistent* $\alpha = ?$ *any other variant* (S8)	$FIC(p,\alpha) = \ln[S^2(p)] + \alpha \sum_{i=0}^{p} v(i,.)$ (S13) $\alpha = 2$ *asymptotically efficient* $\alpha = ?$ *any other criterion*

where the p elements $\hat{a}_i(p)$ constitute the parameter vector $\hat{\boldsymbol{a}}(p)$.

An AR model may be fitted to make a parametrical description of the second order moments of the series. Another reason may be that, faced with questions concerning the future behaviour, a prediction with a model is to be made.

The predictive capacity of a model is expressed by the prediction error, *PE(p)*; a quantity that should not be confused with the residual variance, $S^2(p)$. The latter is a measure of the model fit to the data based upon which the parameters have been estimated and is defined in practice as:

$$S^2(p) = \hat{a}^T(p)\, \hat{R}_{xx}\,(p)\, \hat{a}(p). \tag{3}$$

$\hat{\boldsymbol{R}}_{xx}(p)$ is the estimated pxp covariance matrix of the data $\boldsymbol{x}_n$. The way the entries of this matrix have been estimated depends entirely on the estimation method used. With YW the covariance matrix is hermitian and Toeplitz and furthermore can be shown to be positive definite. With LSF the covariance matrix in (3) is hermitian and positive semidefinite having N-p contributions to its individual elements. In the LSFB method there are $2N$-$2p$ contributions to each of the entries of the matrix, which is hermitian and positive definite in this case. Formula (3) also covers the method of Burg.

The value of $S^2(p)$ decreases for each extra parameter that is included in the model, whereas the forecasting capacities of a model will only improve as long as the extra included parameter is significant. The matrix $\boldsymbol{R}_\infty$ (without the hat!) describes the exact covariance structure of the K-*th* order process that originally generated the data. The *PE(p)* is found by weighting the estimated parameter vector $\hat{\boldsymbol{a}}(p)$ of the fitted model with the pxp submatrix of $\boldsymbol{R}_\infty$ as follows:

$$PE(p) = \hat{a}^T(p)\, R_\infty(p)\, \hat{a}(p). \tag{4}$$

The *PE(p)* of (4) is a quantity that can only be calculated in simulation experiments, since for the calculation of $\boldsymbol{R}_\infty$ the original process must be known. In practice the *PE(p)* can only be estimated, making use of a suitable order selection criterion, like the asymptotical *FPE(p)* (S7) [7] or its finite sample counterpart the *FSC(p)* (S8) [8]).

ELEMENTS OF LARGE SAMPLE THEORY

For the calculation of the bias terms in the parameters use is made of the Yule-Walker equations:

$$\hat{R}_{xx}(p)\, \hat{a}(p) = \hat{r}(p), \tag{5}$$

where $\hat{\boldsymbol{R}}_{xx}(p)$ denotes the doubly symmetric Toeplitz matrix with as first row $R_0 \ldots R_{p-1}$ and $\hat{\boldsymbol{r}}(p)$ is given by $(R_1 \ldots R_p)^T$. The Yule-Walker equations (5) will be used to express the parameters, the residual variance and the prediction error in covariances. These functions are evaluated with a Taylor expansion, yielding expressions of the desired quantities in the second order moments of the measured covariances, which are the fourth order moments of the observations. Three types of bias can now be distinguished:

- Bias caused by subtraction of the estimated mean of the data.
- Triangular bias: The name refers to the bias in the autocovariance estimates $\hat{R}_i$, because it can be seen as if a triangular window has been put over the estimated covariance function. This form of bias appears only in the YW method and arises when $\hat{R}_i$ is divided by N instead of N-i; the expectation of the covariance becomes $(1\text{-}i/N)R_i$. It is the reason that the Yule-Walker method will always produce results that differ from any other estimation method.
- Taylor bias: A bias contribution, derivable by means of the Taylor expansion, which is caused among others by the fact that the expectation of two stochastic variables is not equal to the quotient of their expectations, e.g.:

$$\hat{\rho}_i = \frac{\hat{R}_i}{\hat{R}_0} \mapsto E[\hat{\rho}_i] \approx \frac{E[\hat{R}_i]}{E[\hat{R}_0]} - \frac{\operatorname{cov}[\hat{R}_0, \hat{R}_i]}{E^2[\hat{R}_0]} + \frac{E[\hat{R}_i]\operatorname{var}[\hat{R}_0]}{E^3[\hat{R}_0]}.$$

This gives always a constant bias contribution of $1/N$ for parameters with an even index i, which is also present when the true value of the parameters is zero.

Calculating, as an example, the bias in an AR(2) process the result of the application of the Taylor expansion to the estimates for parameters yields for the Taylor bias:

$$E[\hat{a}_1(2)] \approx a_1 - \frac{a_1}{N}, \quad E[\hat{a}_2(2)] \approx a_2 - \frac{(3a_2 - 1)}{N}.$$

The result for the expectation of $\hat{a}_2(2)$ shows a constant bias

contribution of $1/N$, which is also present when the parameters are zero and is found for all $\hat{a}_i(p)$ with even index i. If the mean of the observations is subtracted, the bias increases with $(1-a_2)/N$ for both parameters. The triangular bias is given by:

$$\text{bias}_{TR}[\hat{a}_1(2)] = \frac{-a_1\{a_1^2-(a_2+1)(3a_2+1)\}}{N\{a_1^2-(a_2+1)^2\}}, \text{bias}_{TR}[\hat{a}_2(2)] = \frac{2a_2(a_2+1)^2}{N\{a_1^2-(a_2+1)^2\}}.$$

The complete bias is found as the sum of the individual elements mentioned above.

WITH PLIM TO TAYLOR RESULTS ?

Expressing $S^2(p)$ and $PE(p)$ as functions of measured covariances and applying a Taylor expansion leads to (S5) and (S6) in a quite straightforward way. Other theoretical derivations have been criticized by Bhansali [9] for their lack of rigorousness. The remainder of this section illustrates the theoretical problems with in some aspects successful derivations. It will be tried to find the Taylor results by means of regression mathematics. The AR process of equation (1) is therefore rewritten in a Least Squares formulation, where the vector of observations $\boldsymbol{x}$, with elements x_1 to x_N is now given by:

$$\boldsymbol{x} = -\boldsymbol{X}\,\boldsymbol{a} + \boldsymbol{\varepsilon}. \qquad (6)$$

If a total of $N+p$ observations is available, the left-hand side vector $\boldsymbol{x}$ of dependent variables becomes $(x_N, x_{N-1}, \ldots, x_1)^T$. It will be assumed that p is much less than N. The most convenient description of the p columns of the matrix of regressors is $(x_{N-1}, x_{N-2}, \ldots, x_0)^T$ for the first column to $(x_{N-p}, x_{N-p-1}, \ldots, x_{1-p})^T$ for the p-th column. Note that the same observations are used at first as the dependent stochastic variables and afterwards as the independent deterministic variables.

An estimate for the parameter vector $\hat{\boldsymbol{a}}$ is:

$$\begin{aligned}\hat{\boldsymbol{a}} &= -(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{x} \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}. \\ &= \boldsymbol{a} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}.\end{aligned}$$

Taking a close look at $\mathbf{X}^T\boldsymbol{\varepsilon}$ it is seen with (1) and (6) that this matrix product is built up of elementary products of the present innovation ε_i with regressors consisting of previous observations x_{i-1} to x_{i-p}, such that the expectation of each individual product equals zero.

Hence

$$E[\,\boldsymbol{X}^T\boldsymbol{\varepsilon}\,] = E[\,\boldsymbol{X}^T\,]E[\,\boldsymbol{\varepsilon}\,] = (\,0\;0\;\ldots\;0\,)^T,$$

with p zeros. The bias should follow from

$$E[\,\hat{\boldsymbol{a}}-\boldsymbol{a}\,] = -E[\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}\,], \qquad (7)$$

but we are not aware of any mathematically sound derivation of the bias of order $1/N$ that is based on this formula.

The residual variance becomes:

$$\begin{aligned}S^2(p) &= [\boldsymbol{x} + \boldsymbol{X}\hat{\boldsymbol{a}}]^T[\boldsymbol{x} + \boldsymbol{X}\hat{\boldsymbol{a}}] / N \\ &= [\boldsymbol{x}^T - \boldsymbol{x}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\,[\boldsymbol{x} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{x}] / N \\ &= \boldsymbol{x}^T[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\,[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\boldsymbol{x} / N \\ &= [\boldsymbol{\varepsilon}^T - \boldsymbol{a}^T\boldsymbol{X}^T]\,[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\,[\boldsymbol{\varepsilon} - \boldsymbol{X}\boldsymbol{a}] / N \\ &= \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} /N - \boldsymbol{\varepsilon}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon} /N \\ &= \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} /N - [\hat{\boldsymbol{a}}-\boldsymbol{a}]^T\boldsymbol{X}^T\boldsymbol{X}\,[\hat{\boldsymbol{a}}-\boldsymbol{a}] /N \\ &= \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} /N - \text{trace}\{\,\boldsymbol{X}^T\boldsymbol{X}\,[\hat{\boldsymbol{a}}-\boldsymbol{a}]\,[\hat{\boldsymbol{a}}-\boldsymbol{a}]^T\,\} /N.\end{aligned}$$

So far only substitutions have been made, no approximations. That will, however, be necessary in the computation of the expectation of the second part of the equation. The trace is taken from Np^2 terms that each consist of a product of two observations and two estimated parameters. The fourth order moments can be written as products of second order moments if the probability density function of the variables involved would be normal. This is certainly not exactly true for estimated parameters: they are at best asymptotically normal. A poorly motivated method to obtain the desired answer is:

$$\begin{aligned}E[\,S^2(p)\,] &\triangleq \sigma_\varepsilon^2 - \text{trace}\{\,E[\,\boldsymbol{X}^T\boldsymbol{X}\,]E[\,[\hat{\boldsymbol{a}}-\boldsymbol{a}][\hat{\boldsymbol{a}}-\boldsymbol{a}]\,\}\,/N \\ &= \sigma_\varepsilon^2 - \text{trace}\{\,E[\,\boldsymbol{X}^T\boldsymbol{X}\,]\,\sigma_\varepsilon^2\,E[\,\boldsymbol{X}^T\boldsymbol{X}\,]^{-1}\,/N \\ &= \sigma_\varepsilon^2\,(\,1 - p/N\,).\end{aligned}$$

The number p follows as the trace of the product of a pxp matrix with its inverse. Both splitting the expectation of fourth order moments in products of second order moments and neglecting those products that are not required in the desired answer have no sufficient mathematical support. It will be extremely difficult to combine mathematically a non-zero outcome for the bias in (7) with the omission of similar terms that is required to obtain the result above for the residual variance.

Since the prediction error can be seen as the fit of $\hat{\boldsymbol{a}}$ to an independent realization $\boldsymbol{y}$ of the same stochastic process, it can be computed with more mathematical reliability because the observations are independent of the estimated parameters (see eqn. 4). The independent equivalents of $\mathbf{x}$ and $\mathbf{X}$ and the innovations $\boldsymbol{\varepsilon}$ are denoted as $\mathbf{y}$, $\mathbf{Y}$ and $\boldsymbol{\eta}$ respectively in (6), with dimension M instead of N. The prediction error for the model with the same order as the generating process, computed from M observations, becomes:

$$\begin{aligned}PE(p) &= E\{\,[\boldsymbol{y} + \boldsymbol{Y}\hat{\boldsymbol{a}}]^T\,[\boldsymbol{y} + \boldsymbol{Y}\hat{\boldsymbol{a}}]\,\} / M \\ &= E\{\,[\boldsymbol{\eta} + \boldsymbol{Y}(\hat{\boldsymbol{a}}-\boldsymbol{a})]^T\,[\boldsymbol{\eta} + \boldsymbol{Y}(\hat{\boldsymbol{a}}-\boldsymbol{a})]\,\} / M \\ &= E\{[\boldsymbol{\eta}^T\boldsymbol{\eta} + \boldsymbol{\eta}^T\boldsymbol{Y}(\hat{\boldsymbol{a}}-\boldsymbol{a}) + (\hat{\boldsymbol{a}}-\boldsymbol{a})^T\boldsymbol{Y}^T\boldsymbol{\eta} + (\hat{\boldsymbol{a}}-\boldsymbol{a})^T\boldsymbol{Y}^T\boldsymbol{Y}(\hat{\boldsymbol{a}}-\boldsymbol{a})]\,\} / M \\ &= \sigma_\varepsilon^2 + 0 + 0 + \text{trace}\{E[\,\boldsymbol{Y}^T\boldsymbol{Y}/M\,]E[\,(\hat{\boldsymbol{a}}-\boldsymbol{a})(\hat{\boldsymbol{a}}-\boldsymbol{a})^T\,]\} \\ &= \sigma_\varepsilon^2 + \text{trace}\{E[\,\boldsymbol{X}^T\boldsymbol{X}/N\,]\,\sigma_\varepsilon^2\,E[\,(\boldsymbol{X}^T\boldsymbol{X})^{-1}\,]\} \\ &= \sigma_\varepsilon^2\,(1 + p/N\,).\end{aligned}$$

The fourth line uses the independence of η and the parameter a. The next line uses the knowledge that X and Y come from a process with the same statistical properties, with N and M rows. In the final step we approximate the expectation of the inverse by the inverse of the expectation, which was already done in the Plim theory [3]. Only this last step lacks mathematical rigour.

Summarizing, we state that the theoretical treatment of the prediction error $PE(p)$ given above is sound and exact up to order $1/N$. It provides the same answer as is found with Taylor. The correct answer for the residual variance can only be found with some mathematically questionable steps. It is not at all possible to find an expression for the bias with this line of reasoning. Summarizing, we could not find modifications of the existing theory that describe the actual finite sample behaviour.

FINITE SAMPLES

Samples are finite when $N<\infty$ and the quotient p/N is greater than, say, 0.1. In finite samples the peculiarities of the method used for the estimation of the parameters become important. The influence of the method on the estimation results has been evaluated by estimating AR parameters in a purely random process. Simulations experiments have been carried out for a number of observations taken between 2 and 100. Most results have been obtained for normally distributed, zero mean innovations ε_n, but a uniform distribution of ε_n yields similar

results. These results are calibration curves for each estimation method. These calibration curves can be described by a set of fairly simple formulae: the $v(i,.)$ of (S9). Given the number of observations N, they do not only describe the influence of the estimation method, but also account for the dependence on the model order i [10]. For all methods $v(0,.) = 1/N$, when the estimated mean of the series is subtracted, otherwise $v(0,.) = 0$. A more detailed motivation for the formulae can be found in [8].

The coefficients $v(i,.)$ correct for the actual number of degrees of freedom that play a role in a given estimation method. A theoretical understanding of the $v(i,.)$ requires to realize that when the value of the quotient i/N differs substantially from zero, the influence of i on the estimation results can no longer be neglected. When N increases, while restricting to the same order i, the quotient i/N may eventually become so small that $v(i,.)$ can be approximated by $1/N$, since $i \ll N$. N becomes simply so large that numerically no difference between $v(i,.)$ and $1/N$ remains. This reasoning is entirely different from the line of reasoning that leads to the asymptotic $1/N$, as a result of restricting the Taylor expansion to the first order.

Another important modification in the Finite Sample Theory is the use of multiplicative formulae, like (S10) and (S11), instead of the additive ones, like (S5) and (S6), in the asymptotic Large Sample Theory. These formulae have been shown to yield better descriptions [8]. It can be illustrated by two figures representing the residual variance $S^2(p)$ and the prediction error $PE(p)$. In Figure 1 the asymptotic description of $S^2(p)$ (S5) is compared to its finite sample counterpart (S10). In Figure 2 the same is done for the $PE(p)$ of (S6) and (S11) respectively. The figures clearly show that simulation results depend heavily on the estimation method that is used. A finite sample description that does not take the estimation method into account is therefore out of the question. For $S^2(p)$ and for $PE(p)$, a single formula describes the different results by using different $v(i,.)$ for each estimation method. The formulae apply to all AR processes above the process order K.

Asymptotical order selection criteria that have information theoretical concepts as their basis, can together be described by the $GIC(p,\alpha)$ (S8). The Finite Sample Theory has been used to rewrite these selection criteria into the $FIC(p,\alpha)$ (S13) (see [8]).

CONCLUSIONS

In this paper a simple and coherent description of the asymptotical Large Sample Theory has been presented. It has been indicated how the statistical expectations of the estimates are found by applying the first order Taylor expansion to the estimates after they have been expressed in estimated covariances. More detailed descriptions are required than the ones thus obtained, especially when the sample sizes become relatively small in comparison with the model order ($p/N > 0.1$). These descriptions are not found by the inclusion of higher order terms in the Taylor expansion. Instead, a different representation has to be used. In the Finite Sample Theory the finite sample variance coefficients provide the descriptions and the necessary corrections for the observed behaviour.

REFERENCES

[1] Broersen, P.M.T. and Wensink, H.E. (1992a). On the theory for autoregressive processes. *Proceedings 1992 International Conference on Acoustics, Speech and Signal Processing*, V-497 - V-500.

[2] Broersen, P.M.T. and Wensink, H.E. (1992b). A framework for autoregressive theory. *Proceedings of EUSIPCO-92*, 775-778.

[3] Mann, H.B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* 11, 173-220.

[4] Kendall, M.G., Stuart, A. and Ord, J.K. (1983). *The Advanced Theory of Statistics, Volume 3* (4th edn). London: Charles Griffin.

[5] Davisson, L.D. (1965). The prediction error of stationary gaussian time series of unknown covariance. *IEEE Trans. on Information Theory* IT-11, 527-532.

[6] Kay, S.M. and Marple, S.L. (1981). Spectrum analysis, a modern perspective. Proc. *IEEE* **69**, 1380-1419.

[7] Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Stat. Math.* 22, 203-217.

[8] Broersen, P.M.T. and Wensink, H.E. (1993). On finite sample theory for autoregressive model order selection. *IEEE Transactions on Signal Processing, vol. 41, no. 1, january 1993*, 194-204.

[9] Bhansali, R.J. (1986). A derivation of the information criteria for selecting autoregressive models. *Adv. Appl. Prob.* 18, 360-387.

[10] Broersen, P.M.T. and Wensink, H.E. (1993). On the penalty factor for AR order selection criteria. *Proceedings of the 14-ième colloque GRETSI.*

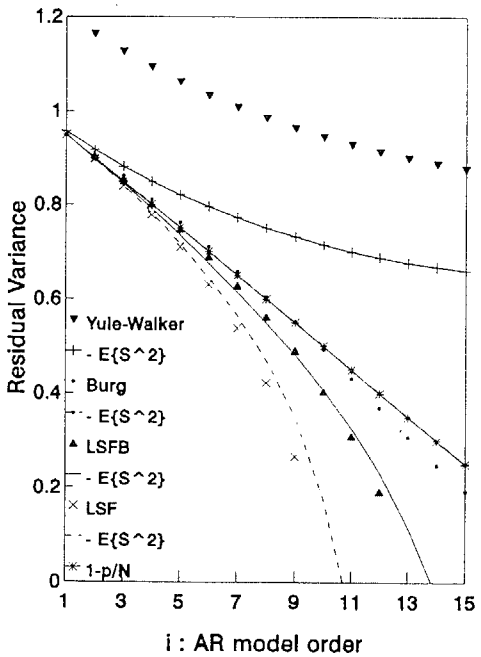

Figure 1: $S^2(p)$ and $E[S^2(p)]$ in the Finite Sample Theory for different estimates based on 20 AR(1) observations

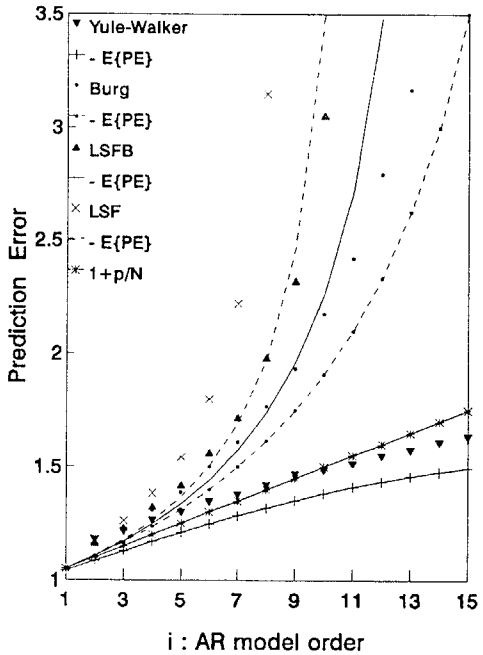

Figure 2: $PE(p)$ and $E[PE(p)]$ in the Finite Sample Theory for different estimates based on 20 AR(1) observations