Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Optimal Calculation of Residuals for ARMAX Models with Applications to Model Verification

Knudsen, Torben

Publication date:
1997

Document Version
Også kaldet Forlagets PDF

Link to publication from Aalborg University

Citation for published version (APA):
Knudsen, T. (1997). *Optimal Calculation of Residuals for ARMAX Models with Applications to Model Verification*.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: May 13, 2024

Optimal Calculation of Residuals for ARMAX Models with Application to Model Verification

T KNUDSEN*

August 1997

Abstract

Residual tests for sufficient model orders are based on the assumption that prediction errors are white when the model is correct. If an ARMAX system has zeros in the MA part which are close to the unit circle, then the standard predictor can have large transients. Even when the correct model is used there will be large correlations in the transient phase. In this case the standard residual tests are therefore not suitable. A new method based on backforecasting is therefore developed. Simulation and analysis shows that the new method gives the right answer where the standard method is misleading.

1 Introduction

Model verification is an important part of system identification. Statistical methods exist to test if the model has too many or too few parameters. These tests are based on the assumption that the optimal one step prediction errors are white noise when the correct model structure and the correct parameters are used.

Using the standard formulas for predicting the output from an ARMAX system, the prediction error are in general not white before a transient phase has passed. This is true even when the correct model and parameters are used. Thus, the statistical methods are based on asymptotic properties which of course are not true for all samples.

The question addressed by this paper is: Does this have any influence on the validity of the tests, and if so, for which models are the influence significant.

The approach is to first take a closer look at the predictor in order to analyze the statistical properties of the prediction error. This will enable us to specify the problem in more detail. Some different solutions are discussed. The standard procedure is then compared to the most promising solution by simulation and analysis. Finally a conclusion is made.

The transient nature of the prediction errors has of course also consequences for the well known prediction error method. Readers are referred to [1, 2] where similar problems concerning parameter estimation are discussed and generalizations needed for general SISO models are presented.

The notation follows [3] and is explained as used. Notice that convergence involving random sequences are convergence in mean square sense.

*Institute of Electronic Systems, Department of Control Engineering, Aalborg University, Fredrik Bajers Vej 7, DK-9220 Aalborg, Denmark

2 ARMAX models and the one step predictor

The ARMAX model can be formulated as (1)–(2) where $\mathrm{ID}(0,\sigma^2)$ is short for independent distributed with mean 0 and variance σ^2. Notice that $A(q)$ and $C(q)$ are monic and that a unit time delay from u to y is assumed for simplicity. Assume also that $C(q)$ is stable.

$$A(q)y(t) = B(q)u(t) + C(q)e(t)\ ,\ e(t) \in \mathrm{ID}(0,\sigma^2) \tag{1}$$

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \\ B(q) &= b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b} \\ C(q) &= 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c} \end{aligned} \tag{2}$$

The optimal one step predictor is given in (3).

$$\hat{y}(t) = \frac{B(q)}{C(q)}u(t) + \frac{C(q)-A(q)}{C(q)}y(t) \tag{3}$$

To calculate the right side of (3) we need measurements of u and y from time $t-1$ and back to the infinite past. In this case, that is the stationary case, the prediction error $\epsilon(t)$ (6) equals the noise $e(t)$ and the predictor is truly optimal. However we do not have measurements back to the infinite past so consequently an initialization procedure has to be chosen.

If the parameter vector, the signal vector and the prediction error are defined as in (4)–(6), the optimal predictor can also be written as (7).

$$\begin{aligned} \theta = (&a_1, a_2, \ldots, a_{n_a}, b_1, b_2, \ldots, b_{n_b} \\ &, c_1, c_2, \ldots, c_{n_c})^T \end{aligned} \tag{4}$$

$$\varphi(t) = (-y(t-1), \ldots, -y(t-n_a) \\ , u(t-1), \ldots, u(t-n_b) \\ , \epsilon(t-1), \ldots, \epsilon(t-n_c))^T \quad (5)$$

$$\epsilon(t) = y(t) - \hat{y}(t) \quad (6)$$

$$\hat{y}(t) = \varphi(t)^T \theta \quad (7)$$

Let us define the time for the first measurement as 1. Then the predictor can be calculated by (6)–(7) for $t \geq t_s$, $t_s = \max(n_a, n_b) + 1$, when it is initialized by (8).

$$\varphi(t_s) = (-y(t_s-1), \ldots, -y(t_s-n_a) \\ , u(t_s-1), \ldots, u(t_s-n_b) \\ , 0, \ldots, 0)^T \Rightarrow \quad (8)$$

$$\epsilon(t_s-1) = \cdots = \epsilon(t_s-n_c) = 0 \quad (9)$$

The idea behind this reasonable procedure is not to start predicting until the measurements of u, y needed in φ are available and at that time to set the missing prediction errors to the *unconditional* mean for e i.e. zero. This initialization procedure is also suggested by [4, p 491] and [3, p 277], and it will be called the direct start (DS).

3 Autocovariance function for the prediction error when using the direct start

Now it is possible to make a theoretical calculation of the autocovariance function for the prediction error. To do this we need a stochastic model for ϵ, that is to say, a model with u and e as inputs.

$$\begin{aligned} \epsilon(t) &= y(t) - \hat{y}(t) \\ &= y(t) - [(1 - A(q))y(t) + B(q)u(t) \\ &\quad + (C(q) - 1)\epsilon(t)] \\ &= A(q)y(t) - B(q)u(t) + (1 - C(q))\epsilon(t) \\ &= C(q)e(t) + (1 - C(q))\epsilon(t) \Rightarrow \\ C(q)\epsilon(t) &= C(q)e(t) \ , \ t \geq t_s \end{aligned} \quad (10)$$

The first prediction error (11) has the variance (12).

$$\epsilon(t_s) = C(q)e(t_s) = \sum_{i=0}^{n_c} c_i e(t_s - i) \Rightarrow \quad (11)$$

$$V(\epsilon(t_s)) = \sigma^2 \sum_{i=0}^{n_c} c_i^2 \quad (12)$$

At this point we notice that:

- All statistical properties for $\epsilon(t)$ are specified by C and σ.
- $\epsilon(t) \to e(t)$ as $t \to \infty$.
- The variance for the first prediction error will in general grow with the degree of $C(q)$. For a first order system $V(\epsilon(t_s)) = \sigma^2(1 + c_1^2) \leq 2\sigma^2$.

State space models are convenient for analysis of time varying properties. Consequently we will rewrite (10) and (9) into spate space form.

First define a new variable z by (13). Notice that z only depends on $e(t)$ for $t \in [t_s - n_c, t_s - 1]$. Therefore $z(t), e(t)$ are uncorrelated when $t \geq t_s$.

$$\begin{aligned} z(t) &= \epsilon(t) - e(t) \\ (10) &\Leftrightarrow C(q)z(t) = 0 \ , \ t \geq t_s \\ (9) &\Rightarrow z(t) = -e(t) \ , \ t_s - n_c \leq t \leq t_s - 1 \end{aligned} \quad (13)$$

The state vector x is defined by (14). And the state space models then become (15)–(16) with the initial conditions(17)–(18).

$$x(t) = (z(t), z(t-1), \ldots, z(t-n_c+1))^T \quad (14)$$

$$x(t) = \Phi x(t-1) \ , \ t \geq t_s \quad (15)$$

$$\epsilon(t) = \Gamma x(t) + e(t) \ , \ t \geq t_s \quad (16)$$

$$\Phi = \begin{bmatrix} -c_1 & -c_2 & -c_3 & \cdots & -c_{n_c} \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$\Gamma = (1, 0, \ldots, 0)$$

$$\begin{aligned} x(t_s-1) &= (z(t_s-1), \ldots, z(t_s-n_c))^T \\ &= -(e(t_s-1), \ldots, e(t_s-n_c))^T \Rightarrow \end{aligned} \quad (17)$$

$$\mathrm{Cov}[x(t_s-1)] = \sigma^2 I \quad (18)$$

From above it follows that $x(t_2)$ and $e(t_1)$ are uncorrelated when $t_2, t_1 \geq t_s$.

Using all this it is not difficult to find the autocovariance function for ϵ.

$$(15) \Rightarrow x(t) = \Phi^{t-(t_s-1)} x(t_s-1) \ , \ t \geq t_s \Rightarrow$$

$$\mathrm{Cov}[x(t)] = \sigma^2 \Phi^{t-(t_s-1)} {\Phi^{t-(t_s-1)}}^T$$

$$\begin{aligned} &\mathrm{Cov}[\epsilon(t_2), \epsilon(t_1)] \\ &= E[\epsilon(t_2)\epsilon(t_1)^T] \\ &= E[(\Gamma x(t_2) + e(t_2))(\Gamma x(t_1) + e(t_1))^T] \\ &= E[(\Gamma \Phi^{t_2-(t_s-1)} x(t_s-1) + e(t_2)) \\ &\quad \times (\Gamma \Phi^{t_1-(t_s-1)} x(t_s-1) + e(t_1))^T] \\ &= \sigma^2 \Gamma \Phi^{t_2-(t_s-1)} {\Phi^{t_1-(t_s-1)}}^T \Gamma^T + E[e(t_2)e(t_1)] \Rightarrow \end{aligned}$$

$$\mathrm{Cov}[\epsilon(t_2),\epsilon(t_1)] = \begin{cases} \sigma^2\Gamma\Phi^{t_2-(t_s-1)}{\Phi^{t_1-(t_s-1)}}^T\Gamma^T & , \ t_2 \neq t_1 \\ \sigma^2\Gamma\Phi^{t_2-(t_s-1)}{\Phi^{t_1-(t_s-1)}}^T\Gamma^T + \sigma^2 & , \ t_2 = t_1 \end{cases}$$

If a new time scale (19) is defined then $\tau = 1 \Leftrightarrow t = t_s$ i.e. the time for the first prediction. This makes the formulas somewhat concise.

$$\tau = t - (t_s - 1) \tag{19}$$

The variance and autocovariance for the prediction error then becomes:

$$V[\epsilon(\tau_1)] = \sigma^2(\Gamma{\Phi^{\tau_1}}{\Phi^{\tau_1}}^T\Gamma^T + 1) \ , \ \tau_1 \geq 1 \tag{20}$$

$$\mathrm{Cov}[\epsilon(\tau_2),\epsilon(\tau_1)] = \begin{cases} \sigma^2\Gamma\Phi^{\tau_2}{\Phi^{\tau_1}}^T\Gamma^T & , \ \tau_2,\tau_1 \geq 1 \ , \ \tau_2 \neq \tau_1 \\ \sigma^2\Gamma\Phi^{\tau_2-\tau_1}{\Phi^{\tau_1}}^T\Gamma^T & , \ \tau_2 > \tau_1 \geq 1 \end{cases} \tag{21}$$

Because (15)–(16) is a state space model for the transfer function model (10) the eigenvalues for Φ equals the zeros for $C(q)$. The eigenvalues are then inside the unit circle, thus $\Phi^\tau \to \underline{0}$ as $\tau \to \infty$.

Based on (20)–(3) we can now conclude:

- $V[\epsilon(\tau)] \to \sigma^2$ as $\tau \to \infty$
- $\mathrm{Cov}[\epsilon(\tau_2),\epsilon(\tau_1)] \to 0$ as τ_1 or τ_2 or $|\tau_2 - \tau_1| \to \infty \ , \ \tau_1 \neq \tau_2$
- The closer the zeros for $C(q)$ are to the unit circle, the slower the convergence becomes. $|\mathrm{Cov}[\epsilon(\tau_2),\epsilon(\tau_1)]|$ and $V[\epsilon(\tau)]$ will be larger than the stationary values at the beginning of the measurements.

4 Consequences for residual tests - DS start

In the previous section we saw that the statistical properties for the prediction error have a transient phase. The analysis also indicated that the transient is largest for system with $C(q)$ of high order and with zeros close to the unit circle. A very important question to be answered now is: Do the transients have any significant impact on the residual tests for ordinary systems or are "pathological" cases needed to show an effect?

To answer this question we will look at three examples. These examples are used throughout the paper, and are therefore described in more detail than necessary at this point where only C and σ are needed. The output error structure (23) is chosen for the examples. This corresponds to the ARMAX structure (24) i.e. $C(q) = A(q)$. In (22) $\mathrm{NID}(0,\sigma^2)$ is short for normal and independent distributed with mean 0 and variance σ^2.

$$e(t) \in \mathrm{NID}(0,\sigma^2) \ , \ \sigma = 0.1 \tag{22}$$

$$y(t) = \frac{B(q)}{A(q)}u(t) + e(t) \Rightarrow \tag{23}$$

$$A(q)y(t) = B(q)u(t) + A(q)e(t) \tag{24}$$

The input is a PRBS signal switching between ± 1. Notice however that the mean time step is five samples [4, Example 5.11]. The N/S ratio is approximately 10%. The first example is the discrete counterpart to a continuous time first order system with bandwidth $\frac{1}{20}$Hz. The second example is the discrete counterpart to a continuous time systems consisting of the series of a second order system with bandwidth $\frac{1}{20}$Hz and damping factor 0.5 and a first order system with bandwidth $\frac{1}{20}$Hz. The sampling time is 1, thus the sampling frequency is around 20 times the bandwidth. This is not at all an unusual system. The third example is similar to the second except for the damping factor which is 0.1 in order to illustrate the situation with zeros closer to the unit circle. Fig. 1 shows gain and poles for the examples, notice that the poles equals the zeros for $C(q)$.

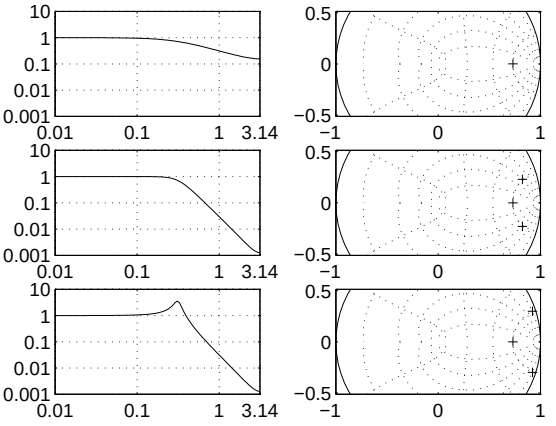

Figure 1: Gain and poles for the examples.

The variance and the autocovariance for the prediction errors are calculated by (20)–(3) and shown in Fig. 2–4 for the examples. It is evident that the transients becomes larger and longer as the zeros for $C(q)$ approaches the unit circle. An important indicator of this is the maximum overshoot for the prediction error variance which are roughly 1.8, 100 and 260 for the first, second and third example respectively.

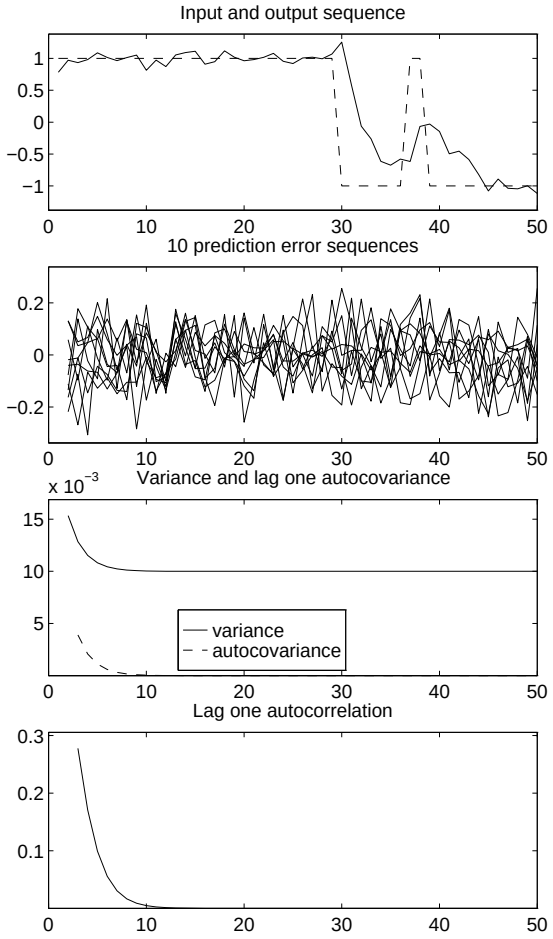

Figure 2: Prediction error properties for the first order example (DS method).

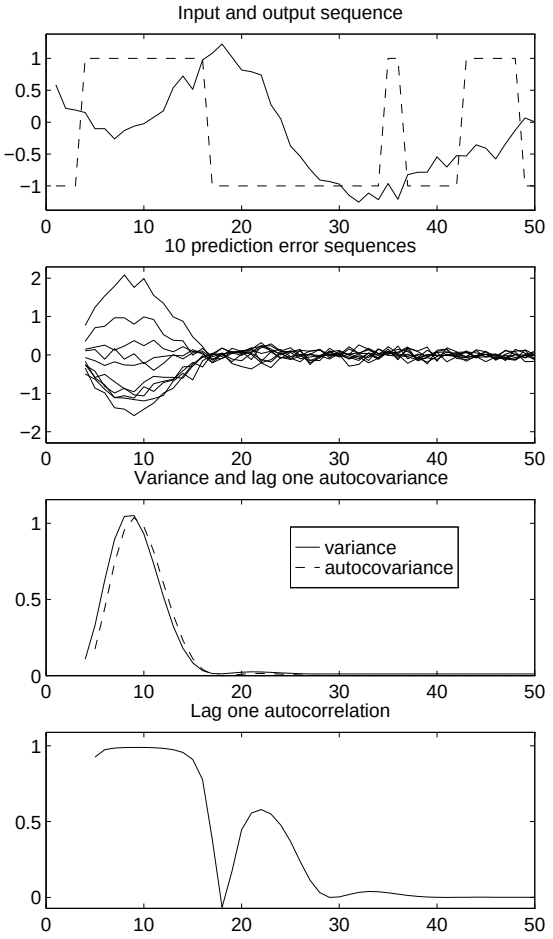

Figure 3: Prediction error properties for the third order system with damping factor 0.5 (DS method).

The autocorrelation test is based on an estimated autocorrelation function. As an example let us calculate the expectation of the estimated lag one autocorrelation for the prediction error.

$$\hat{\rho}_1 = \frac{\widehat{\mathrm{Cov}}[\epsilon(\tau+1),\epsilon(\tau)]}{\widehat{V}[\epsilon(\tau)]}$$

$$\widehat{\mathrm{Cov}}[\epsilon(\tau+1),\epsilon(\tau)] = \frac{1}{n}\sum_{t=1}^{n-1}\epsilon(t+1)\epsilon(t)$$

$$\widehat{V}[\epsilon(\tau)] = \frac{1}{n}\sum_{t=1}^{n}\epsilon(t)^2$$

The expectations for these estimates are:

$$E\{\widehat{V}[\epsilon(\tau)]\} = \frac{1}{n}\sum_{t=1}^{n} V[\epsilon(t)]$$

$$E\{\widehat{\mathrm{Cov}}[\epsilon(\tau+1),\epsilon(\tau)]\} = \frac{1}{n}\sum_{t=1}^{n-1}\mathrm{Cov}[\epsilon(t+1),\epsilon(t)]$$

$$E\{\hat{\rho}_1\} \sim \frac{E\{\widehat{\mathrm{Cov}}[\epsilon(\tau+1),\epsilon(\tau)]\}}{E\{\widehat{V}[\epsilon(\tau)]\}}$$

The resulting expectations are shown in table 1. Rows 1 and 3 are based on the values shown in Fig. 2–3. The deviations from the stationary values are seen to be small for the first order system, but large for the third order system even when using 500 samples. It is important to notice that the expected correlation estimate does not depend on the noise variance, but only on $C(q)$.

The important conclusion to this section is the following. Assume we use the DS and the system parameters for the prediction. Then the variance and the lag one autocovariance for the prediction error has a transient which is significant for an ordinary third order system. The standard autocorrelation test is invalid in this situation, and the S/N ratio has no

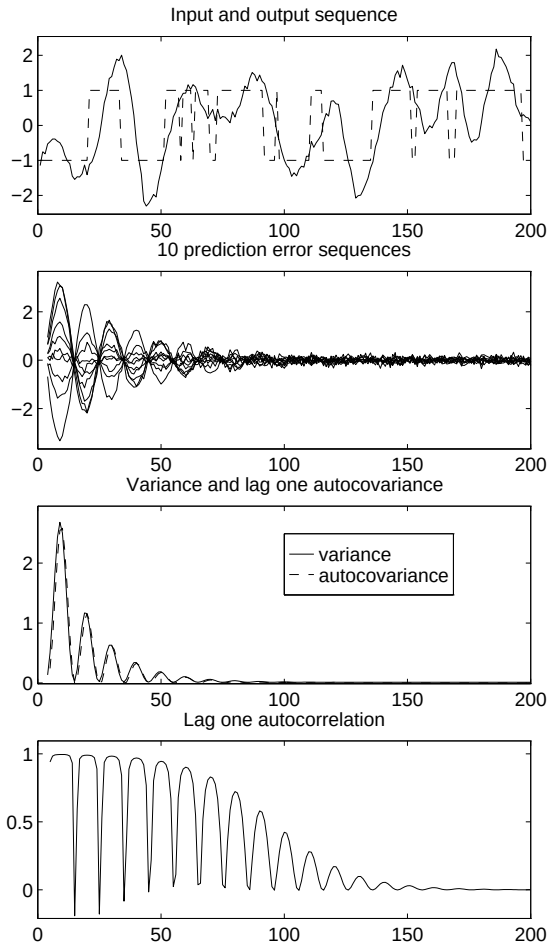

Figure 4: Prediction error properties for the third order system with damping factor 0.1 (DS method).

impact on the validity of the test. If the zeros for $C(q)$ get closer to the unit circle and/or the order of $C(q)$ increases then the transient will also increase. Even a first order system gives problems if the zero is sufficiently close to the unit circle.

5 Solutions

Before turning to specific solutions the problem is reviewed and to main approaches are discussed.

Given an ARMAX structure $A(q), B(q), C(q)$ with $C(q)$ zeros inside the unit circle, a parameter vector θ and some measurements

$$z_1^n = \begin{bmatrix} y_1^n & u_1^n \end{bmatrix}$$
$$y_1^n = \begin{bmatrix} y(1) & \dots & y(n) \end{bmatrix}^T$$
$$u_1^n = \begin{bmatrix} u(1) & \dots & u(n) \end{bmatrix}^T$$

Order	Damp.	# samp.	$E\{\widehat{V}\}$	$E\{\widehat{\mathrm{Cov}}\}$	$E\{\hat{\rho}_1\}$
1	0.5	49	0.01023	0.00017	0.01666
1	0.5	499	0.01002	0.00002	0.00167
3	0.5	47	0.15474	0.14064	0.90889
3	0.5	497	0.02369	0.01330	0.56147
3	0.1	497	0.06594	0.05357	0.81253
Stationary values			0.01	0	0

Table 1: Expected values for estimates of the statistical properties for the prediction error when using the DS method.

The problem is to obtain the corresponding noise sequence $e(t)$ to test it for white noise properties. Further assumptions are avoided to obtain a general solution.

The classical prediction error approach is based on the fact that one step prediction error equals $e(t)$ in the stationary case i.e.

$$\epsilon(t) = y(t) - \hat{y}(t|t-1) \to e(t) \quad \text{as} \quad t \to \infty$$

The notation $\hat{y}(t|t-1)$ emphasizes that $\hat{y}(t)$ is based on past measurements z_1^{t-1}. Consequently $\epsilon(t)$ is based on z_1^t i.e. the measurements from the beginning to the present which are very few in the beginning.

In view of the problems with the transient it would be better to estimate $e(t)$ using all data i.e. like $\mathrm{E}(e(t)|z_1^n)$ because this is the best estimator in the MSE sense.

However, some prediction error based methods are discussed first.

5.1 Discarding the first part of the samples

Simply discard the transient phase from the prediction error sequence. Even through this principle is extremely simple it still require some of the calculations in section 3 to decide on the number of samples to discard. Anyway, this solution is far better than ignoring the problem. It can be recommended in cases with plenty of samples, but it is unsatisfactory with few samples.

5.2 Using a Kalman filter

The ARMAX model (1) can be represented in e.g. companion state space form as follows.

$$x(t+1) = \Phi x(t) + \Gamma u(t) + \Pi e(t) \tag{25}$$
$$y(t) = Hx(t) + e(t) \tag{26}$$
$$\Phi = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix}, \ \Gamma = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{27}$$

$$\Pi = \begin{bmatrix} c_1 - a_1 \\ \vdots \\ c_n - a_n \end{bmatrix} \ , \ H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \tag{28}$$

$$R_1 = \mathrm{Cov}(\Pi e(t)) = \sigma^2 \Pi \, \Pi^T \tag{29}$$

$$R_2 = \mathrm{V}(e(t)) = \sigma^2 \ , \ R_{12} = \sigma^2 \Pi \tag{30}$$

$$\mu_x(0) = \mathrm{E}(x(0)) \ , \ P_x(0) = \mathrm{Cov}(x(0)) \tag{31}$$

Based on (25)–(31) a Kalman filter [5, sec. 11.3] can provide the optimal prediction $\hat{y}(t)$ for $y(t)$ and the time varying prediction error variance (32). The classical residual test requires a stationary error sequence which can be obtained by normalizing with the prediction error standard deviation as in (33).

$$P_y(t) = H P_x(t) H^T + \sigma^2 \tag{32}$$

$$\epsilon(t) = \frac{\hat{y}(t) - y(t)}{\sqrt{P_y(t)}} \tag{33}$$

At this point a serious problem arises which makes it impossible to use the Kalman filter above in the ARMAX case. To obtain the optimal predictions from the beginning it is necessary to use the exact initial conditions (31) which are impossible to obtain because $\mu_x(0)$ depends on the past input which in general is unknown. A number of ad hoc solutions to this problem can be found.

- Use a crude estimate of $\mu_x(0)$ and a correspondingly large estimate of $P_x(0)$.
- Assuming $u(t)$ to be a stochastic process with know properties enables the calculation of stationary values for μ_x and P_x which can be used as initial conditions under the assumption that the system is stationary prior to the measurements, see e.g. [6].
- An alternative to the above approach would be to base the Kalman filter on the state space model for the MA(n_c) auxiliary process $w(t)$ (44) discussed in the next section. This would also avoid the dependence on deterministic past.
- The initial conditions $\mu_x(0)$ can be estimated as the one minimizing the sum of squared prediction errors produced by the Kalman filter using a constant Kalman gain i.e. assuming stationarity. This is similar to the approach suggested for parameter estimation in [4, sect. 12.6].

To summarize on the Kalman filter approaches, they all suffer from one or more disadvantages as: complex calculations, unrealistic assumption and lack of theoretical foundation.

The fundamental drawback for the prediction error approach, including Kalman filtering, still is that $\epsilon(t)$ only is based on z_1^t i.e. part off the know data z_1^n.

5.3 Conditional expectation using all data

When all data are used to estimate the noise the term prediction error is not appropriate, therefore the more general term residual are used.

The relation between $e(t)$ and the measurements is given by (34)–(35).

$$A(q)y(t) - B(q)u(t) = C(q)e(t) \Leftrightarrow \tag{34}$$

$$w(t) = C(q)e(t) \ , \ w(t) = A(q)y(t) - B(q)u(t) \tag{35}$$

In this contents the measurements can therefor be represented by the auxiliary sequence (36) which can be calculated exactly from z_1^n.

$$w_{t_s}^n = \begin{bmatrix} w(t_s) & \ldots & w(n) \end{bmatrix}^T \tag{36}$$

The problem now is to find the conditional expectation (37) where the operator $\check{}$ is introduced for conveniences.

$$\check{e}(t) = \mathrm{E}(e(t)|w_{t_s}^n) \tag{37}$$

The solution is given below where the notation $\{M\}_{ij}$ refers to the element in row i column j in matrix M.

Theorem 1. *Assume that the stochastic part of the ARMAX process is stationary and $e(t) \in NID(0, \sigma^2)$, the conditional expectation for e is then given by*

$$\check{e}_{t_s-n_c}^n = R_{ew} R_w^{-1} w_{t_s}^n \tag{38}$$

where

$$\{R_{ew}\}_{ij} = \begin{cases} \sigma^2 c_{n_c+j-i} & \text{for } 0 \le i-j \le n_c \\ 0 & \text{otherwise} \end{cases}$$

$$\{R_w\}_{ij} = \begin{cases} \sigma^2 r_w(i-j) & \text{for } |i-j| \le n_c \\ 0 & \text{otherwise} \end{cases}$$

$$r_w(k) = \sigma^2 \sum_{i=o}^{n_c-|k|} c_i c_{i+|k|}$$

and

$$E(e_{t_s-n_c}^n - \check{e}_{t_s-n_c}^n) = \begin{bmatrix} 0 & \ldots & 0 \end{bmatrix}^T$$

$$Cov(e_{t_s-n_c}^n - \check{e}_{t_s-n_c}^n) = \sigma^2 I - R_{ew} R_w^{-1} R_{ew}^T$$

$$Cov(\check{e}_{t_s-n_c}^n) = R_{ew} R_w^{-1} R_{ew}^T$$

Remark 1.1. No assumption on the zeros for $C(q)$ is needed, they may be on or even outside the unit circle.

Remark 1.2. If $e(t)$ is not normal distributed, $\check{e}_{t_s-n_c}^n$ (38) may not be the conditional expectation. Consequently, it may not be the best estimate in the mean square sense but it is the best *linear* estimate.

Proof. The dimension of the variables used are listed below.

Variable	dimension
$e^n_{t_s-n_c}$	$n-t_s+n_c+1 \times 1$
$w^n_{t_s}$	$n-t_s+1 \times 1$
R_e	$n-t_s+n_c+1 \times n-t_s+n_c+1$
R_{ew}	$n-t_s+n_c+1 \times n-t_s+1$
R_w	$n-t_s+1 \times n-t_s+1$

The vector (39) has a normal distribution with mean (40) and covariance (41).

$$v = \begin{bmatrix} e^n_{t_s-n_c} \\ w^n_{t_s} \end{bmatrix} \tag{39}$$

$$\mathrm{E}(v) = \begin{bmatrix} 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}^T \tag{40}$$

$$\mathrm{Cov}(v) = \begin{bmatrix} R_e & R_{ew} \\ R_{ew}^T & R_w \end{bmatrix} \tag{41}$$

$$\begin{aligned} R_e &= \mathrm{Cov}(e^n_{t_s-n_c}) = \sigma^2 I \\ R_{ew} &= \mathrm{Cov}(e^n_{t_s-n_c}, w^n_{t_s}) \\ &= \mathrm{E}(e^n_{t_s-n_c} {w^n_{t_s}}^T) \end{aligned} \Rightarrow$$

$$\begin{aligned} \{R_{ew}\}_{ij} &= \mathrm{E}(e(t_s-n_c+i-1)w(t_s+j-1)) \\ &= \mathrm{E}(w(t)e(t-n_c+i-j)) \\ &= \sigma^2 c_{n_c+j-i} \end{aligned} \tag{42}$$

$$\begin{aligned} R_w &= \mathrm{Cov}(w^n_{t_s}) \\ &= \mathrm{E}(w^n_{t_s} {w^n_{t_s}}^T) \end{aligned} \Rightarrow$$

$$\begin{aligned} \{R_w\}_{ij} &= \mathrm{E}(w(t_s+i-1)w(t_s+j-1)) \\ &= \mathrm{E}(w(i)w(j)) \\ &= r_w(i-j) \end{aligned} \tag{43}$$

$$r_w(k) = \mathrm{E}(w(t)w(t+k)) = \sigma^2 \sum_{i=o}^{n_c-|k|} c_i c_{i+|k|}$$

The stationarity is used in (42) and (43).

According to the well known theorem proven in e.g. [7, sec. 7.3], the conditional expectation is given by

$$\begin{aligned} \check{e}^n_{t_s-n_c} &= \mathrm{E}(e^n_{t_s-n_c}|w^n_{t_s}) \\ &= \mathrm{E}(e^n_{t_s-n_c}) + R_{ew}R_w^{-1}(w^n_{t_s} - \mathrm{E}(w^n_{t_s})) \\ &= R_{ew}R_w^{-1}w^n_{t_s} \end{aligned}$$

and the estimation error has zero mean and covariance

$$\mathrm{Cov}(e^n_{t_s-n_c} - \check{e}^n_{t_s-n_c}) = R_e - R_{ew}R_w^{-1}{R_{ew}}^T$$

Finally, the covariance for the residuals is

$$\begin{aligned} \mathrm{Cov}(\check{e}^n_{t_s-n_c}) &= \mathrm{Cov}(R_{ew}R_w^{-1}w^n_{t_s}) \\ &= R_{ew}R_w^{-1}\mathrm{Cov}(w^n_{t_s})R_w^{-1}{R_{ew}}^T \\ &= R_{ew}R_w^{-1}R_wR_w^{-1}{R_{ew}}^T \\ &= R_{ew}R_w^{-1}{R_{ew}}^T \end{aligned}$$

which completes the proof. □

The method above gives the best estimate of $e(t)$. However, it requires a huge amount of computation, especially the inversion of R_w is a problem.

5.4 The backforecasting method

The so called backforecasting (BC) method presented below is an computationally more suitable alternative to the method in theorem 1 because it only uses filtering.

The BC method has been used on ARMA models by Box and Jenkins [8]. Unfortunately this particular method can only be directly applied to ARMAX models if the input $u(t)$ is known back in time, which is not usually the case. Therefore a method for ARMAX models based on the same principles is presented below.

Theorem 2. *Assume that $C(q)$ in an ARMAX system has all zeros inside the unit circle and that $e(t) \in ID(0,\sigma^2)$ then $\check{e}(t)$ calculated by algorithm 1 is an approximation for $\check{e}(t) = E(e(t)|w^n_{t_s})$ with the property*

$$\check{e}(t) \to \check{e}(t) \quad as \quad n \to \infty \ \ \forall\, t \in [t_s - n_c, n]$$

Algorithm 1 (Backforecasting).

1. Calculate $w(t)$ for $t = t_s, t_s+1, \dots, n$.

   $$w(t) = A(q)y(t) - B(q)u(t)$$

2. Calculate $\check{e}_b(t)$ backwards for $t = n, n-1, \dots, t_s$, initialize with $\check{e}_b(n+1), \dots, \check{e}_b(n+n_c) = (0, \dots, 0)$.

   $$\check{e}_b(t) = w(t) - c_1\check{e}_b(t+1) - \cdots - c_{n_c}\check{e}_b(t+n_c)$$

3. Multi step backforecasting of $w(t)$ for $t = t_s-1, t_s-2, \dots, t_s-n_c$, using $\check{e}_b(t) = 0 \ \ \forall\, t \le t_s - 1$.

   $$\breve{w}(t) = \check{e}_b(t) + \cdots + c_{n_c}\check{e}_b(t+n_c)$$

4. Calculate $\check{e}(t)$ for $t = t_s-n_c, t_s-n_c+1, \dots, t_s-1$, using $\check{e}(t) = 0 \ \ \forall\, t \le t_s - n_c - 1$.

   $$\check{e}(t) = \breve{w}(t) - c_1\check{e}(t-1) - \cdots - c_{n_c}\check{e}(t-n_c)$$

5. Calculate the remaining part of $\check{e}(t)$ i.e. for $t = t_s, t_s+1, \dots, n$ either by

   (a) the filter in step 4 with $\breve{w}(t) = w(t) \ \ \forall\, t \ge t_s$

   or

   (b) use $(\check{e}(t_s-1), \dots, \check{e}(t_s-n_c))$ for the missing initial conditions $(\epsilon(t_s-1), \dots \epsilon(t_s-n_c))$ in the usual prediction error formulas (6)–(7), then $\epsilon(t)$ will equal $\check{e}(t)$.

Remark 2.1. Notice that only filtering is used in the algorithm.

Remark 2.2. Using the filter in step 4 for all data makes step five unnecessary. The motivation for step 5 is to show that the algorithm can be separated in step 1–4 which calculates the initial conditions for the first prediction $\hat{y}(t_s)$ and step 5(b) which based on these initial conditions calculates the residuals in the usual prediction error way.

Proof. The key points in this proof are that the sequence $w(t)$ with the forward model (44) is a MA(n_c) process which equally well can be modeled by the backward model (45). The backward model is developed further to show the notation used. Notice also that e and e_b are different sequences.

$$w(t) = C(q)e(t)\ ,\ e(t) \in \mathrm{ID}(0,\sigma^2) \tag{44}$$

$$\begin{aligned} &w(t) = C(q^{-1})e_b(t)\ ,\ e_b(t) \in \mathrm{ID}(0,\sigma^2) \Leftrightarrow \\ &w(t) = (1 + c_1 q + \cdots + c_{n_c} q^{n_c})e_b(t) \\ &\quad = e_b(t) + c_1 e_b(t+1) + \cdots + c_{n_c} e_b(t+n_c) \end{aligned} \tag{45}$$

Taking conditional expectation on both sides of (44) yields (46). Notice that $\breve{}$ denotes conditional expectation with respect to $w_{t_s}^n$ in general.

$$\breve{w}(t) = C(q)\breve{e}(t) \Leftrightarrow \tag{46}$$

$$\breve{e}(t) = \breve{w}(t) - c_1\breve{e}(t-1) - \cdots - c_{n_c}\breve{e}(t-n_c) \tag{47}$$

It follows from (44), i.e. $w(t)$ being an MA(n_c) process, that $e(t-k), w(t)$ are independent for $k \geq n_c + 1$ and that $w(t+k), w(t)$ are independent for $|k| \geq n_c + 1$ which implies (48) and (49) respectively.

$$\breve{e}(t) = \mathrm{E}(e(t)|w_{t_s}^n) = \mathrm{E}(e(t)) = 0\ \ \forall\, t \leq t_s - n_c - 1 \tag{48}$$

$$\breve{w}(t) = \mathrm{E}(w(t)|w_{t_s}^n) = \mathrm{E}(w(t)) = 0\ \ \forall\, t \leq t_s - n_c - 1 \tag{49}$$

It follows from (35) that $w(t)$ is known for $t \in [t_s, n]$ i.e.

$$\breve{w}(t) = \mathrm{E}(w(t)|w_{t_s}^n) = w(t)\ \ \forall\, t \in [t_s, n] \tag{50}$$

To calculate $\breve{e}(t)$ for $t = t_s - n_c, \ldots, n$ by the filter (47) only $\breve{w}(t)$ for $t \in [t_s - n_c, t_s - 1]$ are missing because the rest of $\breve{w}(t)$ is given by (50) and the initial conditions is given by (48).

The values $\breve{w}(t)$ for $t \in [t_s - n_c, t_s - 1]$ can be found by using the backward model (45) for backforecasting. Taking conditional expectation on both sides of (45) yields (51). Thus $\breve{e}_b(t)$ can be calculated by (52) backwards i.e. for $t = n, n-1, \ldots, t_s$.

$$\breve{w}(t) = C(q^{-1})\breve{e}_b(t) \Leftrightarrow \tag{51}$$

$$\breve{e}_b(t) = w(t) - c_1\breve{e}_b(t+1) - \cdots - c_{n_c}\breve{e}_b(t+n_c) \tag{52}$$

Starting this filter for $t = n$ requires the initial conditions $(\breve{e}_b(n+1), \ldots, \breve{e}_b(n+n_c))$ which are unknown. If these are set to zero a slightly different sequence (53) is obtained which is called $\breve{e}_b(t)$. The notation $\breve{}$ are also used for other sequences which are affected by this approximation.

$$\breve{e}_b(t) = w(t) - c_1\breve{e}_b(t+1) - \cdots - c_{n_c}\breve{e}_b(t+n_c) \tag{53}$$

$$,\ (\breve{e}_b(n+1), \ldots, \breve{e}_b(n+n_c)) = (0, \ldots, 0) \tag{54}$$

However, only the n_c first values $\breve{e}_b(t_s), \ldots, \breve{e}_b(t_s + n_c - 1)$ are needed in the following and because all zeros for $C(q)$ are assumed inside the unit circle the effect of initial conditions will vanish if the number of samples is much larger than the length of the impulse response for $\frac{1}{C(q)}$ i.e.

$$\begin{aligned} &(\breve{e}_b(t_s), \ldots, \breve{e}_b(t_s + n_c - 1)) \\ &\qquad \to (\breve{e}_b(t_s), \ldots, \breve{e}_b(t_s + n_c - 1)) \quad \text{as} \quad n \to \infty \end{aligned} \tag{55}$$

Assume for a moment that $\breve{e}_b(t)$ can be calculated. Because $e_b(t) \in \mathrm{ID}(0,\sigma^2)$ it follows that

$$\breve{e}_b(t) = E[e_b(t)|w_{t_s}^n] = E[e_b(t)] = 0\ \ \forall\, t \leq t_s - 1$$

Now $\breve{w}(t)$ for $t = t_s - 1, \ldots, t_s - n_c$ can be calculated (backforecasted) by (51). This first part of $\breve{w}(t)$ together with the last part (50) and the initial conditions (48) are sufficient to calculate $\breve{e}(t)$ by (47) for $t = t_s - n_c, \ldots n$ as was needed.

When the approximation $\breve{e}_b(t)$ is used corresponding approximations $\breve{w}(t)$ $\breve{e}(t)$ are obtained, however (55) implies that

$$\begin{aligned} &\breve{w}(t) \to \breve{w}(t) \quad \text{as} \quad n \to \infty\ \ \forall\, t \in [t_s - n_c, t_s - 1] \Rightarrow \\ &\breve{e}(t) \to \breve{e}(t) \quad \text{as} \quad n \to \infty\ \ \forall\, t \in [t_s - n_c, n] \end{aligned}$$

which completes the proof. □

Theorem 2 above makes it possible to relax the distribution assumption in theorem 1 as follows.

Theorem 3. *The results in theorem 1 holds asymptotically for $n \to \infty$ for any distribution of $e(t)$*

Proof. $\breve{e}(t)$ is calculated by filtering only, therefore it will be a linear function of data. Then theorem 2 implies that the conditional expectation $\breve{e}(t)$, which is the optimal estimate in the MSE sense, tends to a linear function, now $\breve{e}_{t_s-n_c}^n$ is the optimal linear estimate, in the MSE sense, for any distribution of $e(t)$ therefore it also is the conditional expectation in the limit. □

From an application point of view this section can be summarized as follows. If the impulse response for $\frac{1}{C(q)}$ can be assumed to be shorter than the data sequence the computationally efficient BC algorithm should be used to calculate the residuals. If this is not the case e.g. if the zeros for $C(q)$ is on the unit circle the BC method will not work but then the method in theorem 1 can be used. Finally, if the impulse response for $\frac{1}{C(q)}$ is known to be negligible the DS can be used.

6 Consequences for residual tests - BC start

To compare the BC method with the DS method the results from the tree examples described in section 4 are shown below.

The number of samples are chosen sufficiently large to apply the BC method. This method i.e. algorithm 1 is used to calculated the residuals in the second subplots in figure 5–7, clearly no transient are visible here. With the DS method the prediction errors could be calculated only from time t_s and forward but the BC method can also give the n_c values before t_s these are however not shown in the figures.

The statistical properties for the residuals is given (approximately) by theorem 1 and shown in the last two subplots where t_s is indicated by the first tick-mark and vertical dotted line. Clearly there are transients in the statistical properties but they are very small compared to the corresponding ones from the DS method. Larger deviations from the stationary values can only be observed for the first n_c samples which is the reason to exclude them from the calculated sequences.

Based on the above results the expected values for the important estimates can be calculated and are show in table 2. It can be concluded that the BC method succeeds to produce estimates with expected values with a negligible deviation from the theoretical ones. Comparing with table 1 it is seen that this is not at all the cases using the DS method.

Order	Damp.	# samp.	$E\{\hat{V}\}$	$E\{\widehat{\mathrm{Cov}}\}$	$E\{\hat{\rho}_1\}$
1	0.5	49	0.00989	-0.00008	-0.00804
1	0.5	499	0.00999	-0.00001	-0.00078
3	0.5	47	0.00962	-0.00035	-0.03588
3	0.5	497	0.00996	-0.00003	-0.00328
3	0.1	497	0.00996	-0.00004	-0.00387
Stationary values			0.01	0	0

Table 2: Expected values for estimates of the statistical properties for the residual when using the BC method. t_s is marked whit an extra tick-mark.

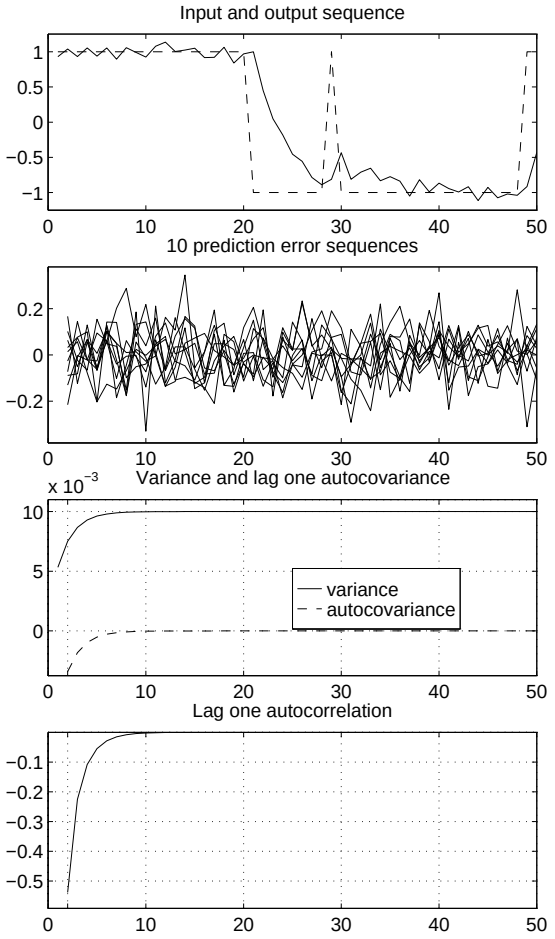

Figure 5: Residual properties for the first order example (BC method).

7 Application to residual test

In this section the third order system with damping factor 0.5 is used to compare the BC procedure with the DS when the prediction errors are applied in residual tests. The number of observations used is 500. It is necessary to know the right parameters and to be able to control the assumptions. For these reasons the comparison is based on simulation. The software used is MATLAB.

As a reference the RESID procedure from the System Identification Toolbox [9] is chosen. The reasons are that this procedure uses some kind of DS and it is written by Lennart Ljung, which gives us every reason to believe that it works well.

According to the analysis the first three white noise samples will specify the transient. A particular 500 sample sequence may or may not show a transient. Of course I have chosen a sequence which gives a transient. Actually the default initial values for the random generators in MATLAB are used.

The middle and bottom sequences in Fig. 8 are based on the correct system parameters and calculated by RESID and the BC algorithm respectively. Clearly only the former gives a transient. The topmost sequence is also calculated by RESID but now with parameters estimated by ARMAX, an parameter estimation procedure from the toolbox. In this case the transient has been reduced. The reason for this is that the ARMAX procedure searches for a minimum of the usual LS criterion which increases dramatically for parameters given a transient as e.g. the system parameters. The resulting estimate will then be biased because it is a compromise between

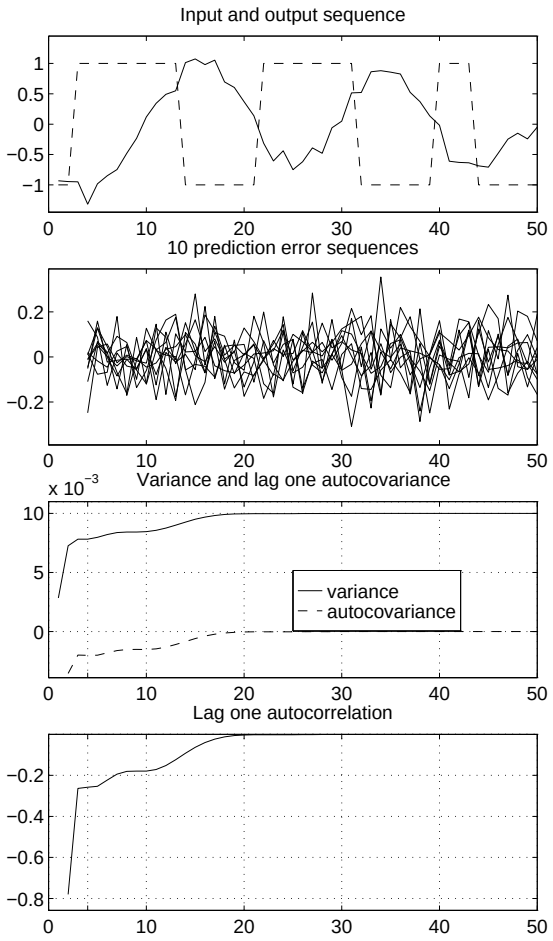

Figure 6: Residual properties for the third order system with damping factor 0.5 (BC method).

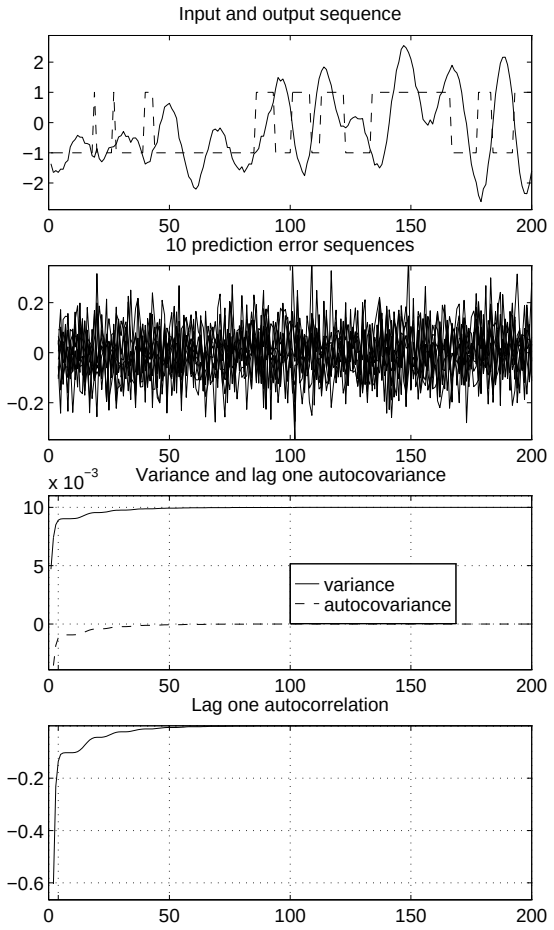

Figure 7: Residual properties for the third order system with damping factor 0.1 (BC method).

minimizing the transient and the stationary part of the sequence [1].

To test if the model is large enough, the RESID procedure graphs the auto- and crosscorrelation estimates with their 99% confidence limits. Figure 9 shows the three autocorrelation tests which correspond to the three sequences in Fig. 8. For the reasons explained the estimated parameters pass the test using the RESID procedure. Using the RESID procedure one would not accept the system parameters because the autocorrelations exceed the confidence limits for lag 1–7. Finally the BC based procedure gives no reason to reject the system parameters.

8 Conclusion

This paper concerns model testing for ARMAX models. The focus is on problems that occur when the MA part has zeros close to the unit circle.

It is shown that the prediction errors resulting from the optimal one step predictor, initialized in the ordinary way, gives large transients even for a quite ordinary third order system. Thus the stationary properties on which the tests are based are not true for all samples.

By analysis it is shown that this results in severe problems for the standard autocorrelation test when the order of the MA part is larger than around two depending on how close the zeros are to the unit circle, even a first order system can give problems. The prediction error transients will also cause problems for other tests as e.g. tests for too many parameters.

One solution to the problem is to use the measurements to estimate the missing initial condition, which will nearly remove the transient. In this paper a method based upon the principle of backforecasting is developed.

Analysis shows that this method is superior to the ordinary method. By simulation this method is compared to the RESID procedure from the system identification toolbox for MATLAB. For the simulation experiment an ordinary third

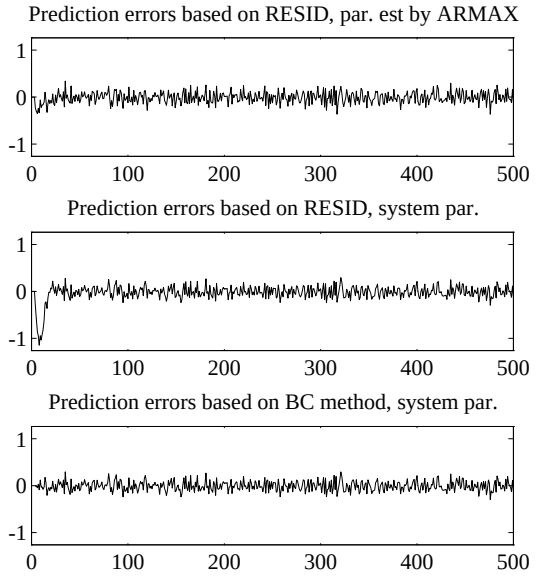

Figure 8: Prediction error/residual sequences from the same input/output sequence.

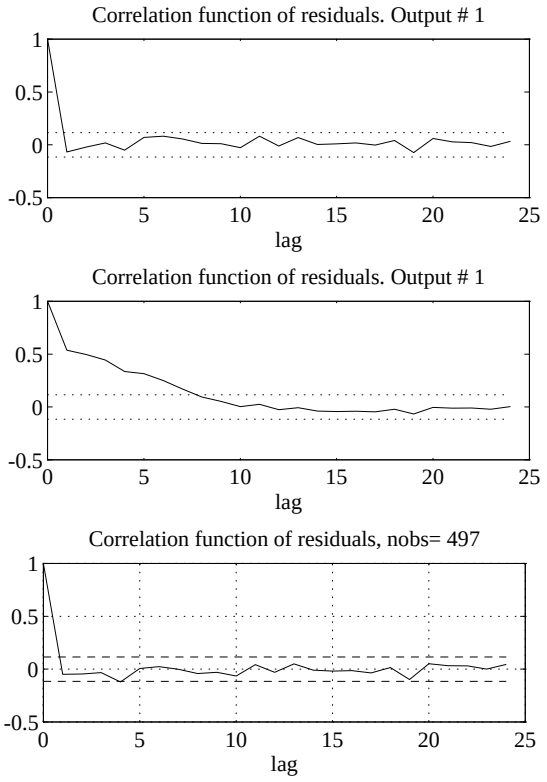

Figure 9: Residual tests

order system is used. Using the RESID procedure one could not accept the system parameters, but using the procedure developed here there was no reason to reject them.

The conclusion is therefore that when working with ARMAX models, especially with MA order larger than 2, one should be careful when using standard test procedures because they can be misleading. It is better to use the procedure developed in this paper. Alternatively if there are plenty of samples it is also possible to simply discard the transient phase from the prediction error sequence.

References

[1] Torben Knudsen. A new method for estimating armax models. In Mogens Blanke and Torsten Söderström, editors, *SYSID'94 10th IFAC Symposium on System Identification, Copenhagen, Denmark, 4-6 July 1994*, volume 2, pages 611–617. Danish Automation Society, July 1994.

[2] Torben Knudsen. The initialization problem in parameter estimation for general siso transfer function models. In Janos J. Gertler, Jr. Jose B. Cruz, and Michael Peshkin, editors, *13th World Congress of IFAC*, volume J, pages 221–226. International Federation of Automatic Control, Elsevier, 1996.

[3] Lennart Ljung. *System Identification, Theory for the User.* Prentice-Hall information and system sciences series. Prentice-Hall, 1987.

[4] Torsten Söderström and Petre Stoica. *System Identification.* Prentice Hall International Series in System and Control Engineering. Prentice Hall, New York, 1989.

[5] Karl J. Åström and Björn Wittenmark. *Computer Controlled Systems: Theory and Design.* Prentice-Hall information and system sciences series. Prentice-Hall, second edition, 1990.

[6] E. J. Hannan and Manfred Deistler. *The Statistical Theory of Linear Systems.* Probability and Mathematical Statistics. John Wiley & Sons, 1988.

[7] Karl J. Åström. *Introduction to Stochastic Control Theory*, volume 70 of *Mathematics in science and engineering.* Academic Press, 1970.

[8] G. E. P. Box and G. M. Jenkins. *Time Series Analysis, Forecasting and Control.* Holden Day, San Francisco, 1976.

[9] Lennart Ljung. *System Identification Toolbox.* The MathWorks, Inc., July 1991.