IFAC Workshop on Adaptation and Learning in Control and Signal Processing, and IFAC Workshop on Periodic Control Systems, Yokohama, Japan, August 30 – September 1, 2004

NON-ASYMPTOTIC CONFIDENCE SETS FOR THE PARAMETERS OF ARMAX MODELS

Marco Campi [*,1] **Erik Weyer** [**,2]

[*] *Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy. Email: campi@ing.unibs.it*
[**] *CSSIP, Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. Email: e.weyer@ee.mu.oz.au*

Abstract: In this paper we consider the problem of constructing confidence sets for the parameters of ARMAX models. Based on subsampling techniques and building on earlier exact finite sample results due to Hartigan, we compute the exact probability that the true parameters belong to certain regions in the parameter space. By intersecting these regions, a confidence set containing the true parameters with guaranteed probability is obtained. All results hold rigorously true for any finite number of data points and no asymptotic theory is involved. Moreover, prior knowledge on the uncertainty affecting the data is reduced to a minimum. A simulation example is provided showing that the method delivers practically useful confidence sets with guaranteed probabilities.

Keywords: System identification, confidence sets, uncertainty evaluation, non-asymptotic theory, ARMAX models.

1. INTRODUCTION

It is widely recognised that a model is of limited use if no certification of its quality is delivered together with the model itself. In principle, a model can be used as if it were the true system provided that it is so accurate that the system-model discrepancy is negligible. However, this is seldom the case, and the model accuracy should be taken into account when the model is used in practice. For the evaluation of model quality one will only have a finite amount of data available. Thus, a sound uncertainty evaluation method must provide results valid when the number of data is finite, and, possibly, small.

Quite often, uncertainty evaluations are based on the asymptotic theory of system identification (e.g. Ljung (1999) or Söderström and Stoica (1988)). It is common experience of theorists and practitioners that this theory - though applied heuristically with a finite number of data points - in many situations delivers sensible results. On the other hand, the correctness of the results is not guaranteed, and contributions (Garatti et. al. (2004)) have appeared that show that the asymptotic theory may as well fail to be reliable in certain situations. Moreover, when the available data is scarce, using asymptotic results makes no sense.

Our earlier finite sample results (e.g. Campi and Weyer (2002) and Weyer and Campi (2002)) were data independent, in the sense that they were uniform with respect to the considered class of data generat-

[1] This research has been supported by MIUR under the project "New Methods for Identification and Adaptive Control of Industrial Plants"
[2] This research has been supported by the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP) under the Cooperative Research Centre scheme funded by The Commonwealth Government

ing systems, and they could essentially be evaluated without any data. Because of the uniformity, it was realised that the results could be quite conservative for the particular system at hand. In this paper we extend the method in Campi and Weyer (2003) for generating guaranteed non-asymptotic confidence regions for the parameters of ARMA models to ARMAX models. The approach uses data generated by the actual system at hand, and hence avoids the problems due to uniformity. Finite sample results using a data based approach has also been developed in Campi et al. (2002), and of course many popular techniques such as bootstrap are data based (e.g. Tjärnström and Ljung (2002)). However, few rigorous finite sample results exists for bootstrap methods.

The methodology developed in this paper does not deliver a nominal model. Instead, it delivers a set of possible models to which the true system belongs with a guaranteed probability. In this respect the methodology has a lot in common with set membership identification, see e.g. Milanese and Vicino (1991), Bai et al. (1996), Giarre' et al. (1997). However, unlike the typical setting in set membership identification we do not need to assume that the disturbances are deterministic or bounded. Loosely speaking, one could view the developed methodology as a stochastic set membership approach to system identification, where the setting we consider is the standard stochastic setting for system identification from e.g. Ljung (1999) or Söderström and Stoica (1988), but where the outcomes are more in line with the outcomes from set membership identification.

The mathematical approach of this paper is inspired by the work of Hartigan (Hartigan (1969)) in the statistical literature. In Hartigan (1969), Hartigan considered the problem of estimating a constant from noisy measurements and introduced the idea that sample estimates based on a certain group theoretical property exhibit special distribution characteristics, valid for a finite number of measurements. The present paper departs from the original work of Hartigan in that we consider more general random sequences (and this allows us to deal with *dynamical* systems).

In the next section we give a simple preview example illustrating the main idea. In section 3 we consider ARMAX models and give the algorithm for the construction of the confidence region and the theoretical results giving the probability that the true parameters belong to the constructed region. Due to space limitations, we have left out the proofs of the theorems, but they are available from the authors on request.

2. A PREVIEW EXAMPLE

In this section, a preview example is given that illustrates the type of results developed. Consider the system

$$y_t + a^0 y_{t-1} = w_t, \tag{1}$$

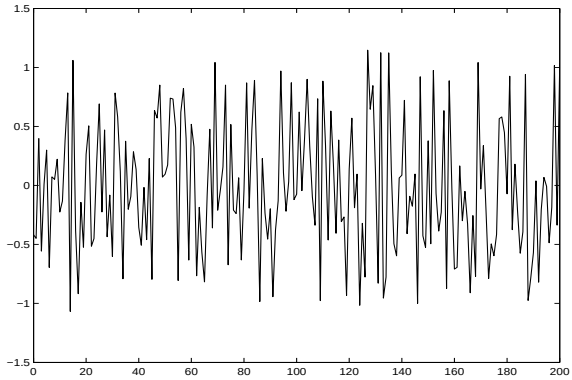

Fig. 1. Data for the preview example

where $a^0 = 0.2$ and $\{w_t\}$ is an independent sequence of random variables uniformly distributed between -1 and 1. 1025 data points were generated according to (1). The first 200 data points are shown in Figure 1. Our goal is to form a confidence region for a^0 from the available data set.

Rewrite the system as a model with generic parameter a:

$$y_t + a y_{t-1} = w_t.$$

The predictor and prediction error associated with the model are

$$\hat{y}_t(a) = -a y_{t-1}, \quad \epsilon_t(a) = y_t - \hat{y}_t(a) = y_t + a y_{t-1}.$$

Next we compute the prediction errors $\epsilon_t(a)$ for $t = 1, \ldots, 1024$ and calculate

$$f_{t-1}(a) = \epsilon_{t-1}(a)\epsilon_t(a), \quad t = 2, \ldots, 1024.$$

Using the $f_{t-1}(a)$'s, we want to form empirical estimates of the correlation $E[\epsilon_{t-1}(a)\epsilon_t(a)]$. Such estimates, however, need to be constructed very carefully. First, we generate a set G of subsets of $I = \{1, \ldots, 1023\}$ which is a group with respect to the symmetric difference, i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G$, if $I_i, I_j \in G$. The generated group is taken from Gordon (1974), and it has 1024 elements and apart from the empty set, each set in G has 512 elements. The sets in G are denoted $I_1, \ldots, I_{1024}$. The incident matrix for a group is a matrix whose (i, j) element is 1 if $j \in I_i$ and zero otherwise. An incident matrix $\bar{R}$ for the 1023 nonempty sets are generated as follows. Let $R(1) = [1]$, and recursively compute ($l = 2, 3, \ldots$)

$$R(2^l - 1) = \begin{bmatrix} R(2^{l-1} - 1) & R(2^{l-1} - 1) & 0 \\ R(2^{l-1} - 1) & J - R(2^{l-1} - 1) & e \\ 0 & e^T & 1 \end{bmatrix}$$

where J and e are, respectively, a matrix and a vector of all ones. Then, $\bar{R} = R(1023)$.

The estimates of the correlation $E[\epsilon_{t-1}(a)\epsilon_t(a)]$ (in fact a re-scaled version as no normalization is present) are then given by

$$g_i(a) = \sum_{k \in I_i} f_k(a), \quad i = 1, \ldots, 1024$$

($g_i(a) = 0$ if $I_i = \emptyset$). A few $g_i(a)$ functions are plotted in Figure 2.

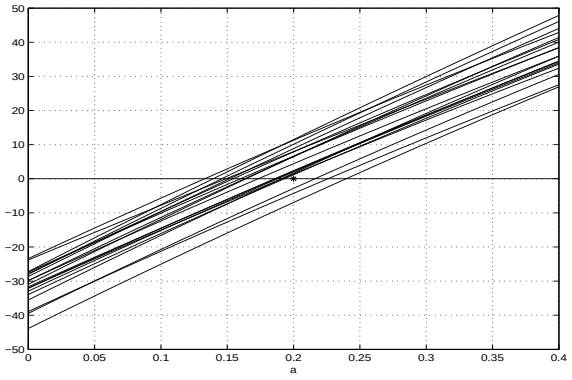

Fig. 2. $g_i(a)$ functions for the preview example

Now, the idea is that for the true a^0, $\epsilon_t(a^0) = w_t$ is white noise and it is very unlikely that all the $g_i(a^0)$ functions but a few will be less than zero or greater than zero. Grounded on this idea, we discard the rightmost and leftmost regions where only 25 functions out of the calculated 1024 are less than zero or greater than zero. The resulting interval [0.12, 0.2425], is the confidence region for a^0. It is a rigorous fact (stated in Theorem 3.1) that this confidence region has probability $1 - 25 * 2/1024 = 0.9512 > 95\%$ to contain the true parameter value a^0.

A verification of the theoretical confidence result was performed by running the same simulation 5000 times. The empirical frequency of a^0 being in the confidence interval was 0.9490, in good agreement with the theoretical result.

3. CONFIDENCE REGIONS FOR ARMAX SYSTEMS

3.1 Data generating system

The ARMAX system that generates the data is given by

$$A^0(z^{-1})y_t = B^0(z^{-1})u_t + C^0(z^{-1})w_t,$$

where

$$A^0(z^{-1}) = 1 + a_1^0 z^{-1} + \cdots + a_n^0 z^{-n}, \quad (2)$$
$$B^0(z^{-1}) = b_1^0 z^{-1} + \cdots + b_m^0 z^{-m}, \quad (3)$$
$$C^0(z^{-1}) = 1 + c_1^0 z^{-1} + \cdots + c_p^0 z^{-p}. \quad (4)$$

It is assumed that the polynomials have no common factors, and that $A^0(z^{-1})$ and $C^0(z^{-1})$ are stable. z^{-1} is the backward shift operator ($z^{-1}u(t) = u(t-1)$). $\{w_t\}$ is a zero-mean independent sequence (noise). No a-priori knowledge of the noise level is assumed. The system operates in open loop, that is $\{w_t\}$ and $\{u_t\}$ are independent. ($\{u_t\}$ can as well be a deterministic signal.)

3.2 Model structure

The model class is $A(z^{-1},\theta)y_t = B(z^{-1},\theta)u_t + C(z^{-1},\theta)w_t$, $\theta \in \Theta$, where $A(z^{-1},\theta), B(z^{-1},\theta)$ and $C(z^{-1},\theta)$ are the same as in (2)-(4) except that a_i^0, b_i^0 and c_i^0 are substituted by a_i, b_i and c_i, $\theta = [a_1,\ldots,a_n,b_1,\ldots,b_m,\ c_1,\ldots c_p]^T$ and $C(z^{-1},\theta)$ is stable for any $\theta \in \Theta$.

3.3 Construction of the confidence set

We start with presenting a procedure for generating certain set Θ_r^ϵ and Θ_s^u, which by Theorem 3.1 below are exact confidence sets for the true parameter θ^0. The final confidence set is obtained by intersecting the Θ_r^ϵ and Θ_s^u sets.

Procedure for the construction of $\Theta_\mathbf{r}^\epsilon$ and $\Theta_\mathbf{s}^\mathbf{u}$

(1) Compute

$$\epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = \frac{A(z^{-1},\theta)}{C(z^{-1},\theta)}y_t - \frac{B(z^{-1},\theta)}{C(z^{-1},\theta)}u_t$$

where $t = 1, 2, \ldots, H$.

(2) For $r = 1,\ldots,p$ and $t = 1+r,\ldots,N+r = H$, compute

$$f_{t-r,r}^\epsilon(\theta) = \epsilon_{t-r}(\theta)\epsilon_t(\theta).$$

(3) For $s = 1,\ldots,n+m$ and $t = 1+s,\ldots,N+s = H$, compute

$$f_{t-s,s}^u(\theta) = u_{t-s}\epsilon_t(\theta).$$

(4) Let $I = \{1,\ldots,N\}$ and consider a collection G of subsets $I_i \subseteq I$, $i = 1,\ldots,M$, forming a group under the symmetric difference operation (i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G$, if $I_i, I_j \in G$). Here M is the number of elements in the group. For $i = 1,\ldots,M$ compute

$$g_{i,r}^\epsilon(\theta) = \sum_{k\in I_i} f_{k,r}^\epsilon(\theta),\ r = 1,\ldots,p,$$
$$g_{i,s}^u(\theta) = \sum_{k\in I_i} f_{k,s}^u(\theta),\ s = 1,\ldots,n+m.$$

(5) Select an integer q in the interval $[1,(M+1)/2)$. For $r = 1,\ldots,p$, find the regions Θ_r^ϵ such that at least q of the $g_{i,r}^\epsilon(\theta)$ functions are bigger than zero and at least q are smaller than zero. Similarly for $s = 1,\ldots,n+m$, find the regions Θ_s^u such that at least q of the $g_{i,s}^u(\theta)$ functions are bigger than zero and at least q are smaller than zero.

Remark 3.1. In the procedure, the group G can be freely selected. Thus, if $I = \{1,2,3,4\}$, a suitable group is $G = \{\{1,2\},\{3,4\},\emptyset,\{1,2,3,4\}\}$; another one is $G = \{\{1\},\{2,3,4\},\emptyset,\{1,2,3,4\}\}$; yet another one is $G =$ all subsets of I. While the theory presented holds for any choice, the quality of the result in the uncertainty region assessment is affected by the choice made. Moreover, the feasible choices are

limited by computational considerations. For example, the set of all subsets cannot normally be chosen as it is a truly large set.

The intuitive idea behind this algorithm is that, for $\theta = \theta^0$, the functions $g^{\epsilon}_{i,r}(\theta)$ and $g^{u}_{i,s}(\theta)$ assume positive or negative value at random ($\epsilon(t, \theta_0)$ is white noise), so that it is unlikely that almost all of them are positive or that almost all of them are negative. Since point 5 in the construction of Θ^{ϵ}_r discards regions where all $g^{\epsilon}_{i,r}(\theta)$'s but a small fraction (q should be taken to be small compared to M, see Theorem 3.1 below) are of the same sign, we expect that $\theta^0 \in \Theta^{\epsilon}_r$ with high probability. Similarly for the construction of Θ^{u}_s. This is put on solid mathematical grounds in the next theorem.

THEOREM 3.1. Assume that variables w_t admit a density (so that $Pr\{w_t = c\} = 0$, for any real c) and that they are symmetrically distributed around zero. Then, the sets Θ^{ϵ}_r and Θ^{u}_s constructed above is such that:

$$Pr\{\theta^0 \in \Theta^{\epsilon}_r\} = 1 - 2q/M, \;\; r = 1, \ldots, p,$$
$$Pr\{\theta^0 \in \Theta^{u}_s\} = 1 - 2q/M, \;\; s = 1, \ldots, n + m.$$

Remark 3.2. When the $\{w_t\}$ process is independent and identically but not symmetrically distributed, we can obtain symmetrically distributed data by considering the difference between two subsequent data points. The noise assumption is mild enough to accommodate a number of situations. In particular, one can describe possible outliers by allowing the noise to take on large values with small probability.

Theorem 3.1 quantifies the probability that θ^0 belongs to the regions Θ^{ϵ}_r and Θ^{u}_s. It holds for any finite N and introduces no conservativeness at all, since such a probability is *exactly* equal to $1 - q/2M$. A good evaluation method must have two properties: the provided region must have guaranteed probability (and this is what Theorem 3.1 delivers); the region must be restricted, and, in particular, it should concentrate around θ^0 as the number of data points increases. We next provide a result that shows that the second property is fulfilled, provided that the input $\{u_t\}$ is white. After the theorem, we discuss this condition and see how it can be relaxed.

THEOREM 3.2. Let $\epsilon_t(\theta) = \frac{A(z^{-1},\theta)}{C(z^{-1},\theta)} y_t - \frac{B(z^{-1},\theta)}{C(z^{-1},\theta)} u_t$ be the prediction error associated with the ARMAX model class. If $\{u_t\}$ is white with spectral density $\Phi_u(\omega) = \lambda_u^2 > 0$, then $\theta = \theta^0 = [a_1^0 \; \cdots \; a_n^0 \; b_1^0 \; \cdots \; b_m^0 \; c_1^0 \; \cdots \; c_p^0]^T$ is the unique solution to the set of equations:

$$E[u_{t-s}\epsilon_t(\theta)] = 0, \quad s = 1, \ldots, n + m, \quad (5)$$
$$E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)] = 0, \quad r = 1, \ldots, p. \quad (6)$$

Theorem 3.2 says that if we simultaneously impose the $n + p + m$ correlation conditions above, then the only solution is the true θ^0. As $N \to \infty$, the functions $g^{\epsilon}_{i,r}(\theta) \to E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)]$ and $g^{u}_{i,s}(\theta) \to E[u_{t-s}(\theta)\epsilon_t(\theta)]$, provided that the number of elements in each set I_i also tends to infinity. (It is easy to construct groups with this property. Construction of good groups has been considered in Gordon (1974).) This means that each of the regions Θ^{ϵ}_r and Θ^{u}_s gets smaller and the intersection of them gives an uncertainty region shrinking around the true parameter θ^0. This leads to the following result

THEOREM 3.3. Let $\Theta^{\epsilon} = \cap_{r=1}^{p} \Theta^{\epsilon}_r$ and $\Theta^{u} = \cap_{s=1}^{n+m} \Theta^{u}_s$. Furthermore, let

$$\hat{\Theta} = \Theta^{\epsilon} \cap \Theta^{u}.$$

Under the assumptions of Theorem 3.1, the set $\hat{\Theta}$ is such that:

$$Pr\{\theta^0 \in \hat{\Theta}\} \geq 1 - 2(n + m + p)q/M. \quad (7)$$

The inequality in (7) is due to that the sets $\{\theta^0 \notin \Theta^{\epsilon}_r\}$, $r = 1, \ldots, p$ and $\{\theta^0 \notin \Theta^{u}_s\}$, $s = 1, \ldots, n + m$ may be overlapping.

Interestingly enough, the conclusion of Theorem 3.2 does not hold true for coloured input sequences, as shown in Campi and Garrati (2003) and demonstrated by the simulation example in section 4. Assuming that $\{u_t\}$ is white is often unrealistic and we next discuss how to remove this assumption.

Suppose that $\{u_t\}$ is prefiltered by a filter $L(z^{-1})$ before it used in point 3 in the construction of Θ^{u}_s, that is point 3 is substituted by

3'. For $s = 1, \ldots, n + m$ and $t = 1 + s, \ldots, N + s = H$, compute

$$f^{u}_{t-s,s}(\theta) = (L(z^{-1})u_{t-s})\epsilon_t(\theta).$$

Then, Theorem 3.1 (and 3.3) remains valid. In fact, we can also allow the filter $L(z^{-1})$ to be dependent on the input signal, that is, it can be constructed from the input signal without affecting the validity of Theorem 3.1. Moreover if the filter $L(z^{-1})$ is appropriately chosen, θ^0 is the unique solution to the correlation equations, as stated in the next theorem.

THEOREM 3.4. Assume $u_t = Q(z^{-1})\nu_t$ with $\{\nu_t\}$ a white wide-sense stationary sequence of random variables with spectral density $\Phi_\nu(\omega) = \lambda_\nu^2 > 0$ and $Q(z^{-1})$ is a rational and stable transfer function. Let $L(z^{-1}) = Q(z^{-1})^{-1}Q(z)^{-1}$, then θ^0 is the unique solution to the set of equations

$$E[(L(z^{-1})u_{t-s})\epsilon_t(\theta)] = 0, \quad s = 1, \ldots, n + m \quad (8)$$
$$E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)] = 0, \quad r = 1, \ldots, p. \quad (9)$$

The fact that the filter $L(z^{-1})$ is unstable should not be too much of a concern since all operations

are performed in batch so that $L(z^{-1})u_{t-s}$ can be computed as a solution having a causal as well as an anti-causal component.

Also note that an imprecise estimation of $Q(z^{-1})$ does not affect the validity of Theorem 3.1, so that the obtained region does have the guaranteed probability of containing the true θ^0. The issue here is the shape of the region, which, if uniqueness is missing, may comprise spurious portions around the solutions of equations (8) and (9) that do not correspond to θ^0, see section 4 for an example. In that example it is also shown that the spurious regions disappear even if the applied filter is only an approximation of $Q(z^{-1})^{-1}Q(z)^{-1}$.

4. SIMULATIONS

We illustrate the method on an ARMAX system with a non-white input signal. In this example there exists parameter values other than the true one which make the expected value of the correlations zero. However, it is demonstrated that the simple filtering procedure discussed above removes the parameter values not corresponding to the true ones from the confidence set.

The true data generating system is given by

$$y_t = b^0 u_{t-1} + w_t + c_1^0 w_{t-1} + c_2^0 w_{t-2} + c_3^0 w_{t-3}, \quad (10)$$

where $u_t = e_t + h^0 e_{t-1}$, and $b^0 = 1$, $h^0 = 0.8$, $c_1^0 = 2.1$, $c_2^0 = 1.46$, $c_3^0 = 0.336$, i.e. $C^0(z^{-1})$ has zeros at $p_1^0 = -0.6$, $p_2^0 = -0.7$ and $p_3^0 = -0.8$. $\{e_t\}$ and $\{w_t\}$ are mutually independent white Gaussian noise sequences, both with variance 1. We consider a full order model, and the prediction error is given by

$$\epsilon_t(\theta) = \frac{1}{C(z^{-1},\theta)} y_t - \frac{bz^{-1}}{C(z^{-1},\theta)} u_t,$$

where $\theta = [b \quad c_1 \quad c_2 \quad c_3]^T$ and $C(z^{-1}) = 1 + c_1 z^{-1} + c_2 z^{-2} + c_3 z^{-3}$. Following Campi and Garatti (2003) it can be shown that there are two parameter values in addition to the true parameter which make $E[u_{t-1}\epsilon_t(\theta)]$ and $E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)]$, $r = 1, 2, 3$, zero, and a numerical search reveals that they are $\theta_1 = [0.9074 \;\; 2.05 \;\; 1.3974 \;\; 0.3179]^T$ and $\theta_2 = [1.0926 \;\; 2.05 \;\; 1.3974 \;\; 0.3179]^T$ in addition to the true parameters $\theta^0 = [1 \;\; 2.1 \;\; 1.46 \;\; 0.336]^T$.

Next we generated 8191+3 data points ($N = 8191$) according to (10), and calculated the variables

$$\begin{aligned} f^u_{t-1,1}(\theta) &= u_{t-1}\epsilon_t(\theta), && t = 2, \ldots, N+1 \\ f^\epsilon_{t-r,r}(\theta) &= \epsilon_{t-r}(\theta)\epsilon_t(\theta), && r = 1, 2, 3, \\ & && t = 1+r, \ldots, N+r \end{aligned}$$

on a plane in the parameter space containing θ^0, θ_1 and θ_2.

The group we used contained $M = 8192$ sets, and the incident matrix was generated in the same way as in the preview example in section 2. Here $n = 0$, $m = 1$ and $p = 3$, and in order to create a 90% confidence set we excluded the regions in the parameter space where 0 belonged to the 102 largest or smallest values of any of the functions

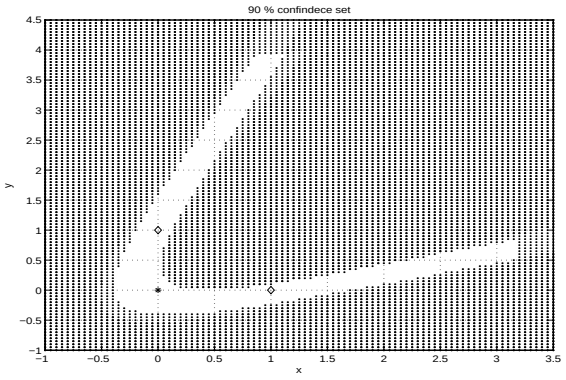

Fig. 3. 90% confidence region. Non-white input signal. $\star$ - true parameter, $\diamond$ - θ_1 and θ_2

$$g^u_{i,1}(\theta) = \sum_{k \in I_i} f^u_{k,1}(\theta), \; i = 1, \ldots, M$$
$$g^\epsilon_{i,r}(\theta) = \sum_{k \in I_i} f^\epsilon_{k,r}(\theta), \; r = 1, \ldots, 3, \; i = 1, \ldots, M.$$

The obtained confidence set has probability at least $1 - 4 * 2 * 102/8192 = 0.9004$, and in Figure 3 we displayed the set in a plane containing the true parameter θ^0 and the two parameters θ_1 and θ_2 which also give zero expected value of the correlations. The parameter values in the plane are given by $\theta^0 + x(\theta_1 - \theta^0) + y(\theta_2 - \theta^0)$ where x and y are the shown coordinates. Hence $(0, 0)$ corresponds to the true parameter, $(1, 0)$ to θ_1 and $(0, 1)$ to θ_2. The obtained confidence set is the blank "boomerang shaped" region, and it contains all three parameters θ^0, θ_1 and θ_2.

Next we generated the filtered input sequence

$$u_t^{\text{fi}} = \left(\frac{1}{1 + 0.9z^{-1}}\right) \cdot \left(\frac{1}{1 + 0.9z}\right) u_t$$

where $\frac{1}{1+0.9z^{-1}}$ is an approximate inverse for $1 + h^0 z^{-1}$. (The filtering was done in a forward and a backward pass thus avoiding stability problems.) u_t was replaced by u_t^{fi} in the calculations of $f^u_{t-1,t}(\theta)$, and the new 90% confidence set is shown in Figure 4. In agreement with the theory, θ_1 and θ_2 no longer belongs to the confidence set which is now concentrated around the true parameters θ^0 (the $(0, 0)$ point).

The filter could also have been estimated from the observed input data. Assuming the model

$$u_t = v_t + h_1 v_{t-1}$$

where v_t is an unobserved zero mean iid sequence, an estimate of h_1 can be obtained using a prediction error method. For the input data in this example we obtained $\hat{h}_1 = 0.7988$. Then using the filtered variable

$$u_t^{\text{fi}} = \left(\frac{1}{1 + \hat{h}_1 z^{-1}}\right) \cdot \left(\frac{1}{1 + \hat{h}_1 z}\right) u_t$$

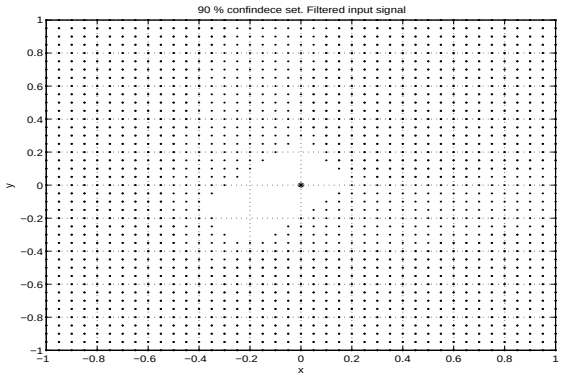

Fig. 4. 90% confidence region. Filtered input signal. $\star$ -true parameter

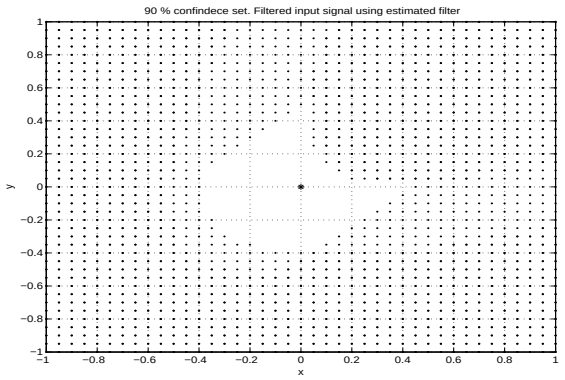

Fig. 5. 90% confidence region. Filtered input signal with estimated filter. $\star$ -true parameter

in the algorithm, we obtain the 90% confidence region shown in Figure 5.

This example shows that for non-white input signals there may be other parameters than the true one which give zero expected value of the correlations, but we can still obtain a confidence set concentrating around the true parameter value by filtering the input sequence. Detailed knowledge of the spectrum of the input signal is not necessary as an approximate filter can produce the desired result as shown in Figures 4 and 5.

5. CONCLUSIONS

In this paper we have derived an algorithm for construction of confidence regions for ARMAX models. The algorithm is based on computing empirical correlation functions using subsamples and discarding regions in the parameter space where only a small fraction of the empirical functions are greater/smaller than zero. Building on finite sample results from Hartigan (1969) we derived bounds, valid for a finite number of data points, on the probability that the true model parameters belong to the constructed region. The developed methodology is grounded on a solid theoretical basis, giving guaranteed probabilities for the true parameter to belong to the constructed set for any finite number of data points, and, as illustrated by the simulation examples, it produces practically useful confidence sets.

REFERENCES

[1] Bai E.W., K.M. Nagpal and R. Tempo (1996). "Bounded-error parameter estimation: noise models and recursive algorithms." *Automatica*, Vol. 32., pp. 985-999.

[2] Campi M.C. and S. Garatti (2003). "Correlation Approach for ARMAX Model Identification: a counterexample on the uniqueness of the asymptotic solution". Internal Report of the University of Brescia.

[3] Campi, M.C., S.K. Ooi and E. Weyer (2002). "Non-asymptotic quality assessment of generalised FIR models". *Proceedings of IEEE Conference on Decision and Control*, pp. 3416-3421, Las Vegas, USA.

[4] Campi, M.C. and E. Weyer (2002). "Finite sample properties of system identification methods". *IEEE Trans. on Automatic Control*, Vol. AC-47, pp. 1329-1334.

[5] Campi M.C., and E. Weyer (2003). "Estimation of confidence regions for the parameters of ARMA models - guaranteed non-asymptotic results." *Proc. 42nd Conf. on Decision and Control*, pp. 6009-6014. Maui, USA, Dec. 2003

[6] S. Garatti, M.C. Campi and S. Bittanti (2004). "Assessing the quality of identified models through the asymptotic theory - When is the result reliable?" To appear in *Automatica*.

[7] Giarre' L., B. Z. Kacewicz and M. Milanese (1997). "Model quality evaluation in set membership identification." *Automatica*, Vol. 33., no. 6, pp. 1133-1139.

[8] Gordon L. (1974). "Completely Separating Groups in Subsampling", *Annals of Statistics* Vol. 2, pp. 572-578.

[9] Hartigan J. A. (1969). "Using Subsample Values as Typical Values", *Journal of American Statistical Association*, Vol. 64, pp. 1303-1317.

[10] Ljung, L. (1999). *System Identification - Theory for the User.* 2nd Ed. Prentice Hall.

[11] Milanese M. and A. Vicino (1991). "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview." *Automatica*, Vol. 27., no. 6, pp. 997-1009.

[12] Söderström, T. and P. Stoica (1988). *System Identification.* Prentice Hall.

[13] Tjärnström F., and L. Ljung (2002). "Using the bootstrap to estimate the variance in the case of undermodelling" *IEEE Transactions on Automatic Control*. Vol. 47, no. 2, pp. 395-398.

[14] Weyer E. and M.C. Campi (2002). "Non-asymptotic confidence ellipsoids for the least squares estimate." *Automatica*, Vol. 38., no. 9, pp. 1539-1547.