Robust Methods for Model Order Estimation

David Hirshberg and Neri Merhav

***Abstract*— Model order estimation is a subject in time series analysis that deals with fitting a parametric model to a vector of observations. This paper focuses on several model order estimators known in the literature and examines their performance under small deviations of the probability distribution of the noise with respect to a nominal distribution assumed in the model. It is demonstrated that the standard estimators suffer from high sensitivity to deviations from the nominal distribution, and a drastic performance degradation is experienced. To overcome this problem, robust estimators that are insensitive to small deviations from the nominal distribution are developed. These estimators are based on a composition between model order estimation methods and robust statistical inference techniques known in the literature. In addition, a new estimator based on a locally best test for weak signals is presented both in nonrobust and robust versions. The proposed robust model order estimators are developed on a heuristic basis, and there is no claim of optimality, but experimental results indicate that they provide significant improvement over the standard estimators.**

I. Introduction

THE problem of estimation the order of a statistical model has been studied in the literature of time-series analysis, information theory, and automatic control. This problem is important in various application areas, like radar and sonar reception, where order estimation is actually detection of the number of targets within the observation sector of the radar or sonar. In science measurement, model order estimation is used, for example, to fit an autoregressive model for sun spot explosions as a function of time, for the prediction of future explosions, or for fitting a model for earthquake measurements. In data compression, the model order is selected in accordance with minimum length considerations. In digital signal processing applications, model order selection can help, for example, in audio noise cancellation and segmentation for image processing.

In this paper, we focus on model order estimators that are defined for general families of probability distributions and, hence, can be applied for a wide range of models. The first attempt to develop such a general model order estimator has been done by Akaike [2]. His estimator, which is called an information criterion (AIC) estimator, proved to be inconsistent in the sense that the error probabilities do not diminish when the number of observations tends to infinity.

Rissanen [14] and Schwarz [15] developed, in different ways, a consistent model order estimator referred to as the minimum description length (MDL) or Bayesian information criterion (BIC). Rissanen obtained this estimator from information-theoretic considerations. The model order in this case is the model that minimizes the description length, i.e., the model that encodes the vector of observations in the most efficient way. Schwarz reached this estimator from a Bayesian approach. He assumed that the model order, as well as the parameters, are random variables and that the model order estimator strives to minimize the overall probability of error. The MDL estimator is the most popular estimator and is widely used in the literature.

Another way to treat the problem of model order estimation is based on hypothesis testing theory. Anderson [1] used hypotheses testing to solve the problem of model order estimation for autoregressive (AR) processes. His approach was to make a sequence of tests between two adjacent model order hypotheses. Anderson proposed the estimated model order to be the highest order that the greater model order hypothesis is declared. Anderson's estimator is difficult to implement and, hence, does not seem to have any later treatment in the literature.

Merhav [12] proposed another type of estimator that was based on hypotheses testing techniques. Merhav extended the known Neyman–Pearson criterion for simple binary hypotheses to the case of model order estimation. The estimator minimizes the probability of underestimation subject to a constraint on the overestimation probability. This estimator is referred to as the Neyman–Pearson criterion (NPC) estimator. For the Markov process model, Merhav *et al.* [13] have shown that if the overestimation probability bound is large enough the estimator is consistent.

The above estimators are based on the assumption that the statistics of the observations produced by the model are known exactly. In the case that the statistics of the observations diverse slightly from the assumed statistics, the estimators suffer from drastic degradation in performance.

Huber introduced methods to overcome the performance degradation problem in parameter estimation [7] and hypotheses testing [6]. Huber's model for uncertainty in the statistics is known as the ϵ-contaminated model. Huber introduced a solution based on the minimax approach that seeks the best solution for the worst-case possible statistics. Comprehensive presentations of the work that has been done in this area can be found in Hample [4], Huber [8], and Kassam and Poor [9]. While the framework of uncertain statistics has been

Manuscript received April 26, 1994; revised June 31, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Geoffrey C. Orsak.

D. Hirshberg was with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel. He is now with the Israeli Ministry of Defense, Rafael, Haifa, Israel.

N. Merhav is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel.

Publisher Item Identifier S 1053-587X(96)02396-1.

studied extensively in the context of parameter estimation and hypotheses testing, to the best of our knowledge, the problem of model order estimation has not yet been treated in this framework.

The purpose of the paper is to introduce the effect of uncertain statistics on known model estimators performance and to suggest new versions of estimators, referred to as the robust estimators, that are more suitable for this case. The robust model order estimators are based on a reasonable composition between the existing model order estimators and the robust statistics techniques. In addition, a new model order estimator both in a robust and a nonrobust version has been developed. This estimator is based on the assumption that model order estimation is a problem of testing close hypotheses. Hence, locally best hypotheses testing techniques, jointly with Anderson's model order estimation approach, is used to construct this model order estimator. These estimators are not claimed to be optimal, but experimental results are provided to demonstrate their good performance.

The outline of the paper is as follows. In Section II, we provide some background. In Section III, robust versions of known model order estimators are developed. In Section IV, we present a new model order estimator in its nonrobust and robust form. Simulation results of the estimators are provided in Section V, and finally, in Section VI, we will summarize our conclusions.

II. Background

A. Model Order Estimation

Let $X = (x(1), x(2), \cdots, x(N))$ be a vector of independent, identically distributed (i.i.d) real-valued random variables governed by a probability density function (pdf) $f(X|\theta^p)$, where θ^p is a vector of parameters of dimension $p, \theta^p = (\theta(1), \theta(2), \cdots, \theta(p)) \in \boldsymbol{R}^p$. A model order estimator $\hat{p}$ is a function of X that returns an integer value. Given a model and an estimator, there are two kinds of error events: underestimation ($\hat{p} < p$) and overestimation ($\hat{p} > p$). P_u and P_o will denote the underestimation and overestimation probabilities, respectively.

The performance of a model order estimator can be judged on the basis of its error probabilities. Typically, there is a tradeoff between these probabilities, i.e., an estimator that has a relatively small P_u will normally have a high P_o and vice versa. A comparison of performance between two estimators can be done by comparing the total error probability $P_u + P_o$ or by comparing the probability of error of one kind for each estimator at a fixed value of the probability of error of the other kind.

The MDL model order estimator is given by

$$\hat{p}_{\mathrm{MDL}} = \arg\min_j \left\{ -\log f(X|\hat{\theta}^j_{\mathrm{ML}}) + \frac{j}{2}\log N \right\} \quad (1)$$

where $\hat{\theta}^j_{\mathrm{ML}}$ is the maximum likelihood estimator under the hypothesis of a j-dimensional parameter vector. The first term in (1) represents the number of bits needed to describe the observations under the best jth order model, and the second term is the optimal number of bits needed to describe the j-dimensional parameter vector.

The NPC model order estimator achieves minimum underestimation probability among all estimators that have overestimation probability that decays as fast as $e^{-\lambda N}$ for a given $\lambda > 0$. This estimator is given by

$$\hat{p}_{\mathrm{NPC}} = \min_j \left\{ j: \frac{1}{N}\log\frac{f(X|\hat{\theta}^{p_0}_{\mathrm{ML}})}{f(X|\hat{\theta}^j_{\mathrm{ML}})} < \lambda \right\} \quad (2)$$

where p_0 is an *a priori* known upper bound on the model order, and λ is a threshold that controls the overestimation decay rate. The estimated model order $\hat{p}$ is the first integer j for which the test accepts the hypothesis that j is the model order and rejects the hypothesis that p_0 is the model order. For the case of Markov processes, Liu and Narayan [10] proved that if the sequence $\lambda = \lambda_N$ satisfies the conditions $\lim_{N\to\infty} \lambda_N = 0$ and $\lim_{N\to\infty} N\lambda_N = \infty$, then the NPC estimator is consistent.

B. Robust Estimation

Huber's model for uncertainty in the statistics is known as the ϵ-contaminated model. According to this model, the distribution function of the observations is a mixture of the known nominal density function $f_N(x)$ and an unknown contaminating density function $f_C(x)$. That is

$$f(x) = (1-\epsilon)f_N(x) + \epsilon f_C(x) \quad (3)$$

where $0 < \epsilon < 1$ is the amount of contamination. For both the parameter estimation and the hypotheses testing problem under the ϵ-contaminated model, Huber introduced a solution based on the minimax approach that seeks the best solution for the worst-case distribution in the ϵ-contaminated distribution family.

In the case of parameter estimation, Huber treated the estimation of a location parameter in the presence of i.i.d noise. The conditional distribution function in this case is

$$f(X|\theta) = \prod_{t=1}^{N} f(x(t) - \theta) \quad (4)$$

where θ is the location parameter. Huber proposed an estimator that has minimal asymptotic variance $V(l, f) \equiv \lim_{N\to\infty} N \cdot \mathrm{Var}(l)$, where l is an estimator, and f is the distribution function of the observations. In the minimax approach, the favored estimator is the one that minimizes the worst-case asymptotic variance $\max_{f\in\mathcal{F}} V(l, f)$, i.e.

$$l_0 = \arg\min_{l\in\mathcal{L}} \max_{f\in\mathcal{F}} V(l, f) \quad (5)$$

where

- l_0 desired robust estimator
- $\mathcal{L}$ family of all possible estimators that meet several regularity conditions
- $\mathcal{F}$ family of ϵ-contaminated distribution functions.

Although the asymptotic variance of the robust estimator is higher than the asymptotic variance of the nonrobust estimator in the presence of nominal noise, the difference is generally small compared with the difference in asymptotic variance when contaminated noise is present. In this case, the asymptotic variance of the nonrobust estimator is much higher than the asymptotic variance of the robust estimator. Huber proved that the robust estimator l_0 is actually a maximum likelihood estimator for a noise density known as the least favorable density function and commonly denoted as f_{LF}. In the case where the nominal distribution is Gaussian with mean zero and unit variance, Huber showed that the least favorable distribution is

$$f_{\mathrm{LF}}(X|\theta) = \left(\frac{1-\epsilon}{\sqrt{2\pi}}\right)^N \exp\left\{-\sum_{t=1}^{N}\rho(x(t)-\theta)\right\} \quad (6)$$

where ρ is given by

$$\rho(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| < K \\ K|x| - \frac{1}{2}K^2 & \text{if } |x| \geq K \end{cases} \quad (7)$$

where K is a function of the amount of contamination ϵ.

As for hypotheses testing, Huber developed a minimax solution as well. In this case, the form of the robust test is similar to the Neyman–Pearson test:

$$T_{\mathrm{NP}} = \prod_{t=1}^{N} \frac{f(x(t)|H_1)}{f(x(t)|H_0)} > k \quad (8)$$

where $f(x(t)|H_0)$ and $f(x(t)|H_1)$ are the conditional densities of each observation under the two hypotheses, and k is the threshold that controls the tradeoff between the false alarm and misdetection probabilities. The robust version of the test limits the effect of the likelihood ratio associated with each observation between lower bound a and upper bound b, i.e.

$$T_{\mathrm{ROBUST}} = \prod_{t=1}^{N} L_a^b\left(\frac{f(x(t)|H_1)}{f(x(t)|H_0)}\right) > k \quad (9)$$

where L_a^b is a limiter function

$$L_a^b(x) = \begin{cases} a & \text{if } x < a \\ x & \text{if } a \leq x \leq b \\ b & \text{if } x > b \end{cases} \quad (10)$$

and the bounds a and b are calculated from the nominal conditional distributions $f(x(t)|H_1)$ and $f(x(t)|H_0)$ and the amount of contamination ϵ.

III. Robust Versions of Known Model Order Estimators

A. Robust MDL Model Order Estimator

Since model order estimators presented in this paper are strongly related to parameter estimation and hypotheses testing, it seems reasonable to apply the techniques used in robust parameter estimation and robust hypotheses testing to robust model order estimation.

For MDL estimators that are based on the likelihood function, we propose robust versions that use the least favorable pdf for parameter estimation $f_{\mathrm{LF}}(X|\theta)$ as defined in (6), instead of the nominal *pdf* $f(X|\theta)$ as the assumed model statistics, that is, we define

$$\hat{p}_{\mathrm{MDL}}^{R} = \arg\min_j \left\{-\log f_{\mathrm{LF}}(X|\hat{\theta}_{\mathrm{LF}}^{j}) + \frac{j}{2}\log N\right\} \quad (11)$$

where the superscript R corresponds to a robust version of the estimator, and $\hat{\theta}_{\mathrm{LF}}^{j}$ is the j-dimensional maximum likelihood estimator under the least favorable distribution assumption.

B. A Robust NPC Model Order Estimator

Since the NPC estimator has the form of a sequential Neyman–Pearson test, it seems that the techniques introduced by Huber [6] for robust hypotheses testing can be applied to construct a robust version of NPC estimator. In this case, the estimator will have the following form:

$$\hat{p} = \min_j \left\{j\colon \frac{1}{N}\sum_{t=1}^{N} L_a^b\left(\log\frac{f(x(t)|\hat{\theta}^{p_0})}{f(x(t)|\hat{\theta}^{j})}\right) < \lambda\right\} \quad (12)$$

where $L_a^b(x)$ is defined in (10). The limiter values depend on the type of the nominal distribution functions and the values of the estimated parameter vectors $\hat{\theta}^{p_0}$ and $\hat{\theta}^{j}$. The parameter vectors $\hat{\theta}^{p_0}$ and $\hat{\theta}^{j}$ can be estimated via nominal maximum likelihood estimation, i.e., $\hat{\theta}^{p_0} = \hat{\theta}_{\mathrm{ML}}^{p_0}$ and $\hat{\theta}^{j} = \hat{\theta}_{\mathrm{ML}}^{j}$ or by using robust parameter estimation, i.e., $\hat{\theta}^{p_0} = \hat{\theta}_{\mathrm{LF}}^{p_0}$ and $\hat{\theta}^{j} = \hat{\theta}_{\mathrm{LF}}^{j}$. The distribution function used for the likelihood ratio test (LRT) could be either the nominal distribution f_N or the least favorable distribution f_{LF}. Simulation results, under the harmonic signal model, for all these possible combinations of the NPC estimator that use robust hypotheses testing techniques reveal poorer performance than the robust version of the NPC estimator, which uses the robust parameter estimation without any use of robust hypotheses testing techniques, that is

$$\hat{p}_{\mathrm{NPC}}^{R} = \min_j \left\{j\colon \frac{1}{N}\log\frac{f_{\mathrm{LF}}(X|\hat{\theta}_{\mathrm{LF}}^{p_0})}{f_{\mathrm{LF}}(X|\hat{\theta}_{\mathrm{LF}}^{j})} < \lambda\right\}. \quad (13)$$

The explanation of the failure of the robust NPC version that uses robust hypotheses testing techniques is that Huber's solution for hypotheses testing is applicable only when the distance between the two hypotheses is large enough. Because of the inherent nested structure of the model order estimation problem, the robust hypotheses testing techniques are not suitable. Martin and Schwartz [11] have faced the same problem in the implementation of a robust version of a signal detector at low signal-to-noise ratios (SNR's). The solution that Martin and Schwartz suggested for this case is to use the locally best hypotheses testing. This technique provides the best performance when the hypotheses are close to each other.

IV. Locally Best Model Order Estimators

Since model order estimation using hypotheses testing techniques deals with close hypotheses, locally best hypotheses testing techniques can be applied. In this section, we develop a new model order estimator based on locally best hypotheses testing techniques. A robust version of this estimator is given as well.

A. Background

Let $H_0: \theta = \theta_0, H_1: \theta = \theta_1$ be two hypotheses where θ_0, θ_1 are nonempty disjoint subsets of the parameter space Θ. A binary test $T(X)$ is a function that divides the sample space Ω into two regions (ω, ω^c), where

$$T(X) = \begin{cases} 1 & \text{if } X \in \omega \\ 0 & \text{if } X \in \omega^c. \end{cases} \tag{14}$$

If $T(X)$ returns the value 0, we say that the test $T(X)$ accepts the hypothesis H_0. Otherwise, the test accepts the hypothesis H_1. A simple hypotheses testing problem occurs when each one of the two hypotheses is associated with one parameter value, i.e., $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$. A composite hypotheses testing problem is the case where at least one of the hypotheses consists of more then one element in the parameter space. A two-sided composite hypotheses testing problem is one where the two hypotheses are $H_0: \theta = \theta_0, H_1: \theta \neq \theta_0$. A locally best test is a test that is optimal only for values of θ in some neighborhood of θ_0. The locally best tests are based on the continuity and the differentiability of the power function of the test defined as

$$\beta_T(\theta) = E_\theta[T(X)]. \tag{15}$$

The value of the power function in θ_0, i.e., $\beta_T(\theta_0)$, is called the size of the test. and it is actually the probability of incorrect decision when H_0 is true. This probability is known also as false alarm probability in signal detection application. A test $T_0(X)$ with size α will be a locally best test if the derivative of its power function in θ_0 is equal zero, i.e.

$$\beta'_{T_0}(\theta_0) = 0 \tag{16}$$

and for all other tests $T(X)$

$$\beta''_{T_0}(\theta_0) < \beta''(\theta_0) \tag{17}$$

where $\beta''_{T_0}(\theta_0)$ is the second derivative at θ_0. The locally best test $T_0(X)$ for the two-sided hypotheses testing problem can be found in Ferguson [3] and has the following form:

$$T_0(X) = \begin{cases} 1 & \text{if } A(X) > B(X) \\ \gamma(X) & \text{if } A(X) = B(X) \\ 0 & \text{if } A(X) < B(X) \end{cases} \tag{18}$$

where

$$A(X) = \frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\Big|_{\theta=\theta_0} + \left(\frac{\partial}{\partial\theta} \log f(X|\theta)\Big|_{\theta=\theta_0}\right)^2$$
$$B(X) = k_1 + k_2 \frac{\partial}{\partial\theta} \log f(X|\theta)\Big|_{\theta=\theta_0}$$

k_1, k_2 being positive constants that comply with (16) and set the size of the test, and $0 \leq \gamma(X) \leq 1$ is an arbitrary measurable function.

B. The Estimator

Similarly to Anderson [1], we propose to perform the model order estimation as a sequence of the two-sided hypotheses tests T_j. Each test T_j is a two-sided test that decides between the hypothesis $H_0: \theta(j) = 0$ and the hypothesis $H_1: \theta(j) \neq 0$, where $\theta(j)$ is the jth element of the parameter vector θ. For $j > p$, we have $\theta(j) = 0$, and therefore, we expect T_j to return 0. For $j \leq p$, we have $\theta(j) \neq 0$; therefore, we expect T_j to return 1. The model order estimator will have the form

$$\hat{p} = \max_j \{j: T_j(X) = 1\}. \tag{19}$$

In each step j, we set the first $j-1$ elements of the parameter vector to be the maximum likelihood estimator, i.e., $\hat{\theta}_{\mathrm{ML}}^{j-1}$ and the jth element of the vector $\theta(j)$ to be the free parameter for the hypotheses testing. By setting T_j to be a locally best test as defined in (18), assigning $\theta = \theta(j)$ and $\theta_0 = 0$, we get the following form for the locally best test T_j:

$$T_j(X) = \begin{cases} 1 & \text{if } A_j(X) > B_j(X) \\ \gamma(X) & \text{if } A_j(X) = B_j(X) \\ 0 & \text{if } A_j(X) < B_j(X) \end{cases} \tag{20}$$

where

$$A_j(X) = \frac{\partial^2}{(\partial\theta(j))^2} \log f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\Big|_{\theta(j)=0} + \left(\frac{\partial}{\partial\theta(j)} \log f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\Big|_{\theta(j)=0}\right)^2$$
$$B_j(X) = k_1 + k_2 \frac{\partial}{\partial\theta(j)} \log f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\Big|_{\theta(j)=0};$$
$$k_1 > 0, k_2 > 0.$$

k_1, k_2 values should satisfy (16), which can be expressed as

$$\beta'_{T_0}(\theta(j)) = \int_\Omega \Pr\{T_j(X)\} \frac{\partial}{\partial\theta(j)} f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\, dX = 0. \tag{21}$$

If $f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))$ is a symmetric function around $\theta(j)$ and, hence, $(\partial/\partial\theta(j)) f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))$ is antisymmetric around $\theta(j)$, then $\Pr\{T_j(X)\}$, which is symmetric around $\theta(j)$, is a sufficient condition for (21) to exist. When $N \to \infty$, $\hat{\theta}_{\mathrm{ML}}^j \xrightarrow{\text{a.s.}} \theta^j$, and hence, $\Pr\{T_j(X)\}$ is symmetric around $\theta(j)$ if $k_2 = 0$. The locally best test model order estimator is reduced, in this case, to the following form:

$$\hat{p}_{\mathrm{LBTC}} = \max_j \left\{ j: \frac{\partial^2}{(\partial\theta(j))^2} \log f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\Big|_{\theta(j)=0} + \left(\frac{\partial}{\partial\theta(j)} \log f(X|(\hat{\theta}_{\mathrm{ML}}^{j-1}, \theta(j)))\Big|_{\theta(j)=0}\right)^2 > k_1 \right\}. \tag{22}$$

We denote this estimator with locally best test criterion (LBTC). This estimator differs from the other estimators in the fact that it uses information on the derivative of $f(X|\theta)$ as well.

Although the estimator seems to be quite complicated, in the general case, it becomes very simple and intuitively appealing in some specific models. For example, in the case of a model where the observations are given by linear combinations of deterministic orthonormal signals in presence of additive white, nominally Gaussian noise, i.e.

$$x(t) = \sum_{j=1}^{p} \theta(j) S_t^j + w(t); \qquad t = 1, 2, \cdots, N \tag{23}$$

where

θ^p parameter vector
S_t^j orthonormal signal family
$w(t)$ white Gaussian noise

the LBTC estimator is given by

$$\hat{p}_{\mathrm{LBTC}} = \max_j \{j \colon |\hat{\theta}_{\mathrm{ML}}^j(j)| > k\} \tag{24}$$

where $\hat{\theta}_{\mathrm{ML}}^j(j)$ is the maximum likelihood estimator for the j-element of the parameter vector.

C. The Robust Version

The robust version of the LBTC estimator, as our all other robust model order estimators, is based on robust parameter estimation. The pdf of the observations for the robust estimator is assumed to be the least favorable pdf for parameter estimation:

$$\hat{p}_{\mathrm{LBTC}}^R = \max_j \left\{ j \colon \frac{\partial^2}{(\partial\theta(j))^2} \log f_{\mathrm{LF}}(X|\hat{\theta}_{\mathrm{LF}}^{j-1}, \theta(j))) \Big|_{\theta(j)=0} + \left(\frac{\partial}{\partial\theta(j)} \log f_{\mathrm{LF}}(X|\hat{\theta}_{\mathrm{LF}}^{j-1}, \theta(j))) \Big|_{\theta(j)=0} \right)^2 > k_1 \right\} \tag{25}$$

where $\hat{\theta}_{\mathrm{LF}}^{j-1}$ is a $(j-1)$-dimensional robust parameter estimator.

V. Experimental Results

The performance of the various estimators have been evaluated by using a Monte Carlo simulation technique for the harmonic signal model. The harmonic signal model is the following:

$$x(t) = \sum_{j=1}^{l} a(j) \sin(j\omega_0 t + \phi(j)) + w(t) \qquad t = 1, 2, \cdots, N \tag{26}$$

where l is the model order, $a^l = (a(1), a(2), \cdots, a(l))$ is the vector of amplitude parameters, $\phi^l = (\phi(1), \phi(2), \cdots, \phi(l))$ is vector of phase parameters, $w(t)$ is a white Gaussian noise with mean zero and variance σ^2, and ω_0 is a known fundamental frequency of the harmonic signal family. The parameter vector $\theta = (a^l, \phi^l, \sigma^2)$ has $2l + 1$ elements in this case. The phase ϕ^l and the amplitude a^p can be treaded as a model with $2l$ amplitude parameters of sine-cosine pairs for each harmonic of the fundamental frequency. Without essential loss of generality, we chose to simulate the model for the case where the phase parameter vector is known and equals to zero. For this model, the conditional pdf is

$$f(X|a^l, \sigma) = \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp \cdot \left\{ -\frac{1}{2\sigma^2} \left[x(t) - \sum_{k=1}^{l} a(k) \sin(k\omega_0 t) \right]^2 \right\} \tag{27}$$

By substituting (27) in each of the above defined estimators and by using simple mathematical manipulations, one can get the following expressions of the various model order estimators under the harmonic signals model

Nonrobust Model Order Estimators:

$$\hat{p}_{\mathrm{MDL}} = \arg\min_j \left\{ \log \hat{\sigma}_{\mathrm{ML}}^j + \frac{j}{2N} \log N \right\} \tag{28}$$

$$\hat{p}_{\mathrm{NPC}} = \min_j \{ j \colon \log \hat{\sigma}_{\mathrm{ML}}^j - \log \hat{\sigma}_{\mathrm{ML}}^{l_0} < \lambda \} \tag{29}$$

$$\hat{p}_{\mathrm{LBTC}} = \max_j \{ |\hat{a}_{\mathrm{ML}}(j)| > k \} \tag{30}$$

where l_0 is an *a priori* known upper bound on the model order, and

$$\hat{\sigma}_{\mathrm{ML}}^j = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(x(t) - \sum_{k=1}^{j} \hat{a}_{\mathrm{ML}}(k) \sin(k\omega_0 t) \right)^2} \tag{31}$$

$$\hat{a}_{\mathrm{ML}}(j) \approx \frac{\sum_{t=1}^{N} x(t) \sin(j\omega_0 t)}{\sum_{t=1}^{N} x(t) \sin^2(j\omega_0 t)} \tag{32}$$

where N is assumed an integer multiple of $2\pi/\omega_0$.

Robust Model Order Estimators: The robust version of the above estimators are as follows.

$$\hat{p}_{\mathrm{MDL}}^R = \arg\min_j \left\{ -\log f_{\mathrm{LF}}(X|\hat{a}_{\mathrm{LF}}^j, \hat{\sigma}_{\mathrm{LF}}^j) + \frac{j}{2N} \log N \right\} \tag{33}$$

$$\hat{p}_{\mathrm{NPC}}^R = \min_j \{ j \colon \log f_{\mathrm{LF}}(X|\hat{a}_{\mathrm{LF}}^j \hat{\sigma}_{\mathrm{LF}}^j) - \log f_{\mathrm{LF}}(X|\hat{a}_{\mathrm{LF}}^{l_0}, \hat{\sigma}_{\mathrm{LF}}^{l_0} < \lambda \} \tag{34}$$

$$\hat{p}_{\mathrm{LBTC}}^R = \max_j \left\{ j \colon -\sum_{t=1}^{N} \rho'' \left(x(t) - \sum_{k=1}^{j-1} \hat{a}_{\mathrm{LF}}^j(k) S_t^k \right) S_t^k + \left[\sum_{t=1}^{N} \rho' \left(x(t) - \sum_{k=1}^{j-1} \hat{a}_{\mathrm{LF}}^j(k) S_t^k \right) S_t^k \right]^2 > k \right\} \tag{35}$$

where

$$f_{\mathrm{LF}}(X|\hat{a}_{\mathrm{LF}}^{j}) = \left(\frac{1-\epsilon}{\sqrt{2\pi}\hat{\sigma}_{\mathrm{LF}}}\right)^{N} \exp \cdot \left\{ -\frac{1}{\hat{\sigma}_{\mathrm{LF}}^{2}} \sum_{t=1}^{N} \rho\left(x(t) - \sum_{k=1}^{j} a_{\mathrm{LF}}^{k} S_{t}^{k} \right) \right\} \tag{36}$$

$$\hat{\sigma}_{\mathrm{LF}}^{j} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(x(t) - \sum_{k=1}^{j}\hat{a}_{\mathrm{LF}}^{j}(k)\sin(k\omega_0 t)\right)^{2}} \tag{37}$$

$$S_t^k = \sin(k\omega_0 t) \tag{38}$$

$$\hat{a}_{\mathrm{LF}}^{j} = \arg\min_{a^j} \left\{ \sum_{t=1}^{N} \rho\left(\sum_{k=1}^{j} a^{j}(k) S_t^k \right) \right\} \tag{39}$$

and $\hat{a}_{\mathrm{LF}}^{j}(k)$ is the k element of $\hat{a}_{\mathrm{LF}}^{j}$.

All these estimators have been examined under various conditions in order to assess their robustness and to investigate the effect of the contaminating pdf, the variance of the contaminating pdf, and the number of observations. Unless specified otherwise, the simulation conditions for each test was as follows:

- The true model order is $l = 2$.
- The amplitude parameter vector is $a = (0.3, 0.3)$.
- The maximum model order is $l_0 = 5$.
- The noise $w(t)$ is distributed by ϵ-contaminated noise with $\epsilon = 0.01$, nominal Gaussian pdf with mean zero, unit standard deviation, and contaminating Laplace pdf with mean zero and standard deviation equal to 100.
- The number of observations is $N = 400$.
- The number of experiments in each test is 1000.

Most of the results are displayed on a (P_u, P_o) plane graph. The performance of the MDL estimators are presented as a point on the graph. The NPC and LBTC estimators are presented by a curve constructed from points depending on the free thresholds λ and k, respectively. The grid in these figures contains, in addition to P_u and P_o, curve lines of $P_u + P_o = C$, where C is a constant.

A. The Effect of the Contaminating pdf

In the following test, we investigate the effect of the type of contamination on the performance of the estimators. The noise generation in this test has been done by applying the ϵ-contamination model with Gaussian, zero mean, and unit variance nominal noise and uniform, Gaussian, Laplace, Rayleigh, and Cauchy types of contamination.

In all cases, except for the Cauchy distribution, the contaminating noise distribution has been with mean zero and variance 100. In the case of Cauchy distribution, where the variance is infinite, we chose to scale the Cauchy distribution in a way that the symmetric tail cut version of the distribution, containing 99% of the distribution mass, has variance 100. "No contamination" is the case where $\epsilon = 0$.

In Fig. 1, the nonrobust and robust MDL estimators performance as a function of the contamination type are shown. The

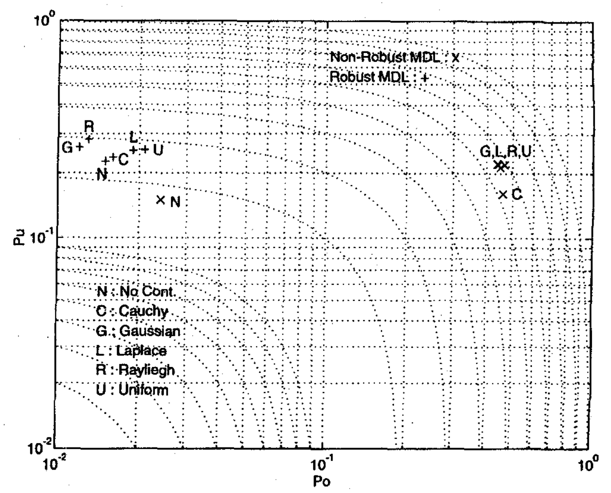


Fig. 1. Effect of contamination type on MDL, nonrobust, and robust estimators.

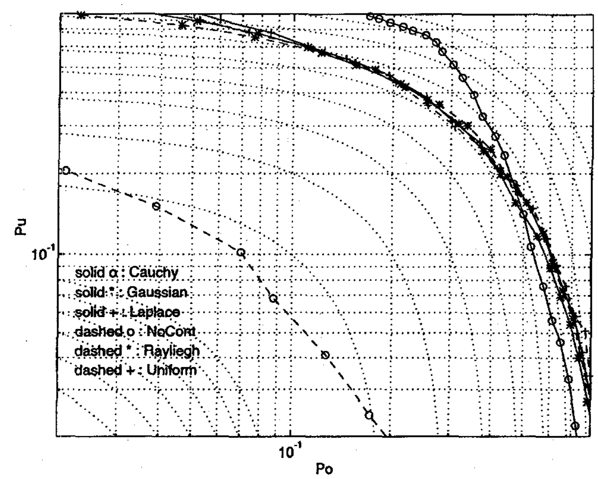


Fig. 2. Effect of contamination type on nonrobust NPC estimator.

nonrobust and robust NPC estimators performance are shown in Figs. 2 and 3, respectively. In Figs. 4 and 5, the nonrobust LBTC and the robust LBTC performance are shown.

In all the nonrobust versions of the model order estimators, we observe a dramatic degradation in performance between where the no contamination case and all other cases, where contamination exists, is observed. For the MDL estimator, the degradation is from ~0.17 to ~0.6. For NPC and LBTC estimators, the degradation depends on the threshold parameters, and a degradation from 0.15 to 0.2 to 0.6–0.7 is observed. As mentioned above, those drastic changes in performance are observed when the amount of contamination ϵ is merely as small as 0.01. On the other hand, all the robust versions of the estimators are insensitive to the contamination type, and the performance is similar for cases of both contaminated noise and noncontaminated noise conditions. The difference in performance between the robust and nonrobust versions of the same estimator type in the presence of optimal conditions, i.e., when no contamination is present, is as small as 0.01–0.04 in

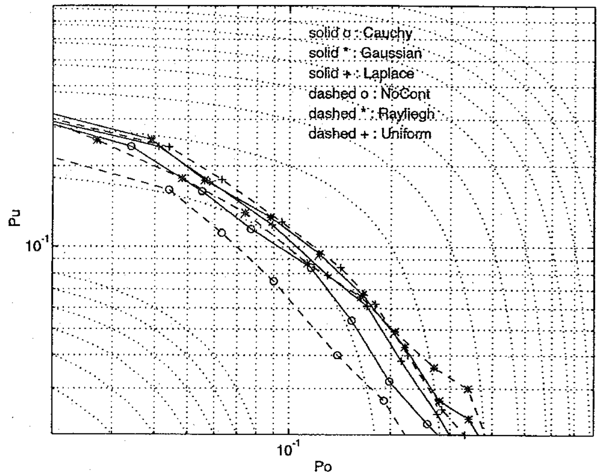


Fig. 3. Effect of contamination type on robust NPC esdtimator.

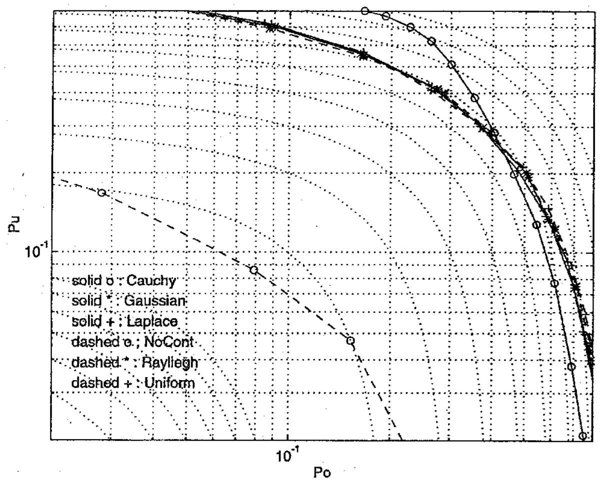


Fig. 4. Effect of contamination type on nonrobust LBTC estimator.

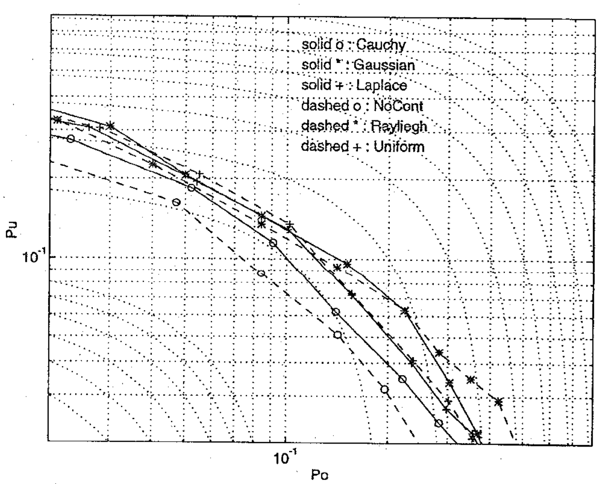


Fig. 5. Effect of contamination type on robust LBTC estimator.

the total probability error, depending on the estimator types. The type of contamination is not so important, and the perfor-

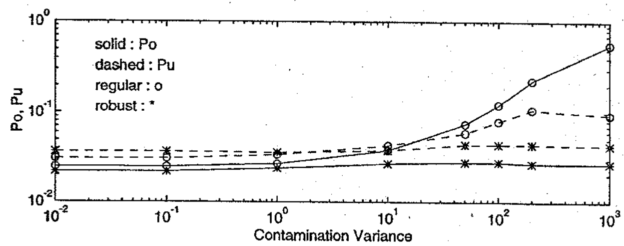


Fig. 6. Effect of contamination variance on MDL nonrobust and robust estimators.

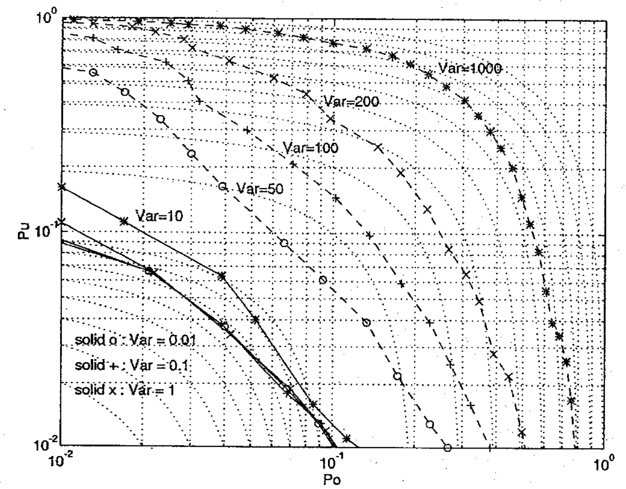


Fig. 7. Effect of contamination variance on nonrobust NPC estimator.

mance is similar for all contamination types both for nonrobust and robust versions of the estimators. As an exception we have the Cauchy contamination, which implies a somewhat different performance in the case of nonrobust estimators. However, for the robust estimators, the performance, in the case of the Cauchy contamination type, is similar to the performance observed for the other types of contamination. The last result could be explained by the fact that the robust versions of the estimator are less sensitive to the existence of contamination and, hence, to the contamination type as well.

B. The Effect of the Variance of the Contamination

In this experiment, we investigate how the variance of the contaminating noise effects the performance of the estimators. The contaminating noise is generated from the Laplace distribution whose variance takes the values of 0.01, 0.1, 1, 10, 50, 100, 200, and 1000.

In Fig. 6, the nonrobust and robust MDL estimators performances are shown. The nonrobust NPC and robust NPC estimators performances are shown in Figs. 7 and 8, respectively. The results for the nonrobust and robust LBTC estimators are similar to these of NPC estimator and will not be presented here.

The contamination variance has a strong effect on the performance of the nonrobust estimators. When the contaminating noise variance is smaller than the nominal distribution variance, no real effect on the performance is detected. However,

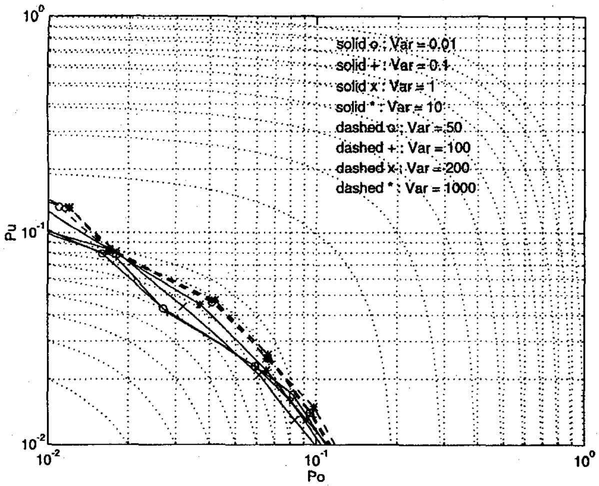


Fig. 8. Effect of contamination type on nonrobust NPC estimator.

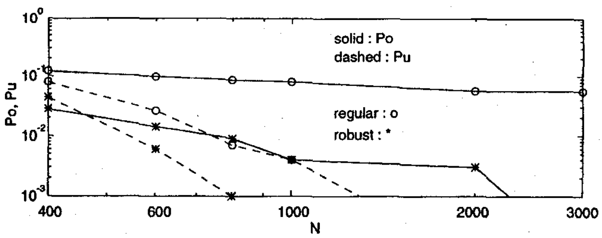


Fig. 9. Effect of the number of observations on MDL nonrobust and robust estimators.

when the variance of the contaminating distribution is of the same order of magnitude as the variance of the nominal distribution or greater, the performance of the nonrobust estimators starts to degrade drastically. The effect of the variance on the robust versions of the estimators is very moderate as opposed to the nonrobust versions. This behavior is expected since increasing the contaminated variance beyond the order of magnitude of the nominal variance produces outliers. As the contaminated variance becomes larger, the outliers become stronger, and their effect on the nonrobust estimator is larger. On the other hand, for the robust estimators, the effect of the outliers is limited; hence, they yield a moderate effect on the performance. In Fig. 6, one can clearly see the moderate performance degradation observed when the contaminating variance becomes larger then the nominal variance. Morever, the degradation in performance for the robust estimators is bounded and does not increase further when the contaminated variance continues to grow.

C. The Effect of the Number of Observations

One of the most interesting features of the estimators is the dependency of the estimator's performance on the number of observations. In this experiment, we investigate the performance of the various estimators as a function of the number of the observations.

In Fig. 9, the nonrobust and robust MDL estimators performances as a function of the number of observations are shown.

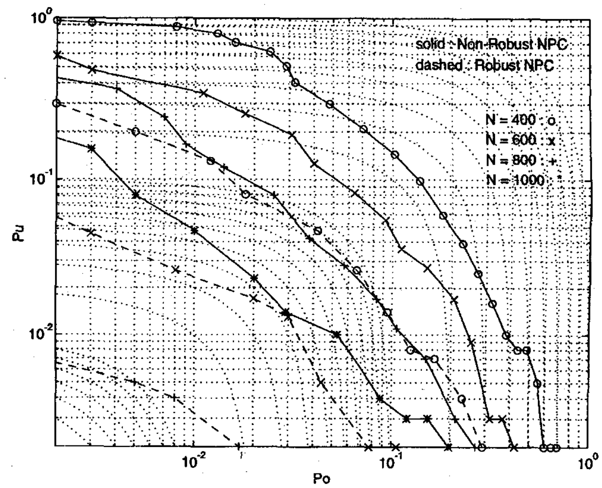


Fig. 10. Effect of the number of observations on nonrobust and robust NPC estimators.

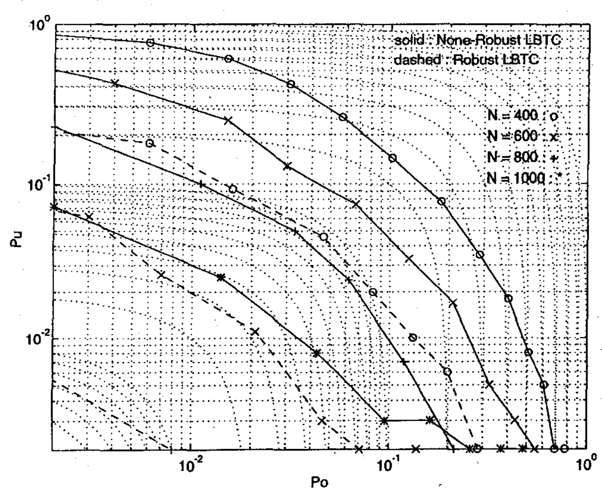


Fig. 11. Effect of the number of observations on nonrobust and robust LBTC estimators.

The performance of the nonrobust and robust NPC estimators are shown in Fig. 10. In Fig. 11, the nonrobust and robust LBTC estimators performances are shown.

For all estimators, it is shown that the probability of both type of errors decreases as N grows. This is important in particular for the LBTC estimator since no theoretical results of consistency have yet been provided. Another interesting phenomenon that can be seen in Fig. 9 is that the overestimation and underestimation error probabilities decay faster for the robust versions of the MDL estimator than for the nonrobust estimator. This implies that one may gain more from using the robust estimator instead of the nonrobust estimator when the number of observations grows.

D. Comparison

In Fig. 12, the performance of all three estimators in their robust versions are shown. Each of these estimator has its working point or curve for $N = 400, 600$, and 800. From

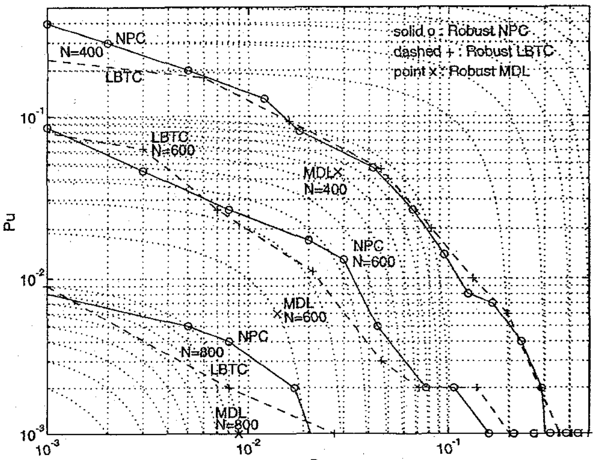


Fig. 12. Comparison between robust model order estimators in uncertainty conditions.

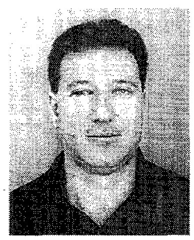
the figure, we can see that MDL estimator has the best performance, and LBTC and NPC are somewhat behind. However, in some other conditions that had been simulated, like when the vector of parameters has a decreasing shape of parameters values, the performance of NPC and LBTC estimators are similar and, in some cases, are even slightly better then the MDL estimators.

VI. Conclusion

Model order estimators, like other statistical inferences, suffer from degradation in performance when the noise distribution is not nominal. Even small deviations from the nominal statistics can yield dramatic performance degradation. In this case, it is very useful to imply robust model order estimators. Several robust versions of model order estimators have been presented in this paper. Among these estimators, the robust version of the MDL estimator seems to be the best choice in general, although in some circumstances, one might consider using the NPC or LBTC estimators as well. Further study can be done to enlarge the theoretical basis of the LBTC estimator and to search for optimal robust estimators in the minimax sense.

References

[1] T. W. Anderson, "Determination of the order of dependence in normally distributed time series," in *Time Series Analysis* (M. Rosenblatt, Ed.). New York: Wiley, 1963, pp. 425–446.
[2] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
[3] T. S. Ferguson, *Mathematical Statistics—A Decision Theoretic Approach.* New York: Academic, 1967.
[4] F. R. Hample, Robust Statistics. New York: Wiley, 1986.
[5] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregretion," *J. Roy. Statist. Soc.*, Ser. B, vol. 41, no. 2, pp. 190–195, 1979.
[6] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, no. 4, pp. 1753–1758, 1965.
[7] ——, "Robust estimation of location parameter," *Ann. Math. Statist.*, vol. 35, no. 4, pp. 73–101, 1964.
[8] ——, *Robust Statistics.* New York: Wiley, 1981.
[9] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, 1985.
[10] C. Liu and P. Narayan, "Order estimation and sequential universal data compression of hidden markov source by the method of mixtures," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1167–1180, July 1994.
[11] R. D. Martin and S. C. Schwartz, "Robust detection of known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50–56, Jan. 1971.
[12] N. Merhav, "The estimation of the model order in exponential families," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1109–1113, Sept. 1989.
[13] N. Merhav, M. Gutman, and J. Ziv, "on the estimation of the order of a Markov chain and universal data compression," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1014–1019, Sept. 1989.
[14] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[15] G. Schwarz, "Estimation the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
[16] M. Wax and T. Kailath, "Detection of signal by the information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.



David Hirshberg was born in Jerusalem, Israel in 1963. He received the B.Sc. and M.Sc. degree in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1986 and 1993, respectively.

From 1986 to 1990, he served in the Israel Defense Force, where he was involved in digital hardware and signal processing system design. During 1991–1995, he was with IBM Science and Technology, Haifa, Israel, working on VLSI design and verification. Since May 1995, he has been with RAFAEL, Israel where he has conducted reserach on communication systems. His research interests are estimation theory, communication, and digital signal processing.

Neri Merhav was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc. and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1982 to 1985, he was a research associate with the Israel IBM Scientific Center in Haifa, where he developed algorithms for speech coding, speech synthesis, and adaptive filtering of speech signals in array sensors. From 1988 to 1990, he was with the Speech Research Department of AT&T Bell Laboratories, Murray Hill, NJ, USA, where he investigated and developed algorithms for speech recognition. He is currently with the Electrical Engineering Department of the Technion and his research interests are statistical signal processing, information theory and statistical communications.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society.