ACCURATE ARMA MODELS WITH DURBIN's SECOND METHOD

P.M.T. Broersen

Department of Applied Physics, Delft University of Technology
P.O.Box 5046, 2600 GA Delft, The Netherlands

ABSTRACT

Long intermediate AR models are used in Durbin's algorithms for ARMA estimation. The order of that long AR model is infinite in the asymptotical theory, but very high AR orders are known to give inaccurate ARMA models in practice. A theoretical derivation is given for two different finite AR orders, as a function of the sample size. The first is the AR order optimal for *prediction* with a purely autoregressive model. The second theoretical AR order is higher and applies if the previously estimated AR *parameters* are used for estimating the MA parameters in Durbin's second, iterative, ARMA method. A Sliding Window (SW) algorithm is presented that selects good long AR orders for data of unknown processes. With a proper choice of the AR order, the accuracy of Durbin's second method approaches the Cramér-Rao bound for the integrated spectrum and the quality remains excellent if less observations are available.

1. INTRODUCTION

Spectral estimation is an important topic in signal processing [1]. Parametric or time series techniques are serious competitors for the non-parametric methods, that use tapered and windowed Fourier transforms [2]. Several solutions have been given for the estimation of AutoRegressive (AR) models, but much less attention has been given to Moving Average (MA) models [2]. Durbin [3] has used long AR models in MA estimation and this method can produce accurate estimates if the order of that AR model is chosen well [4]. The combined ARMA models have different types of solutions [5]. Extended Yule-Walker methods are computationally attractive, but the accuracy may be poor [2]. Maximum likelihood estimation of ARMA parameters from given data is a non-linear problem, which can give rise to difficulties in practice. The results for large samples are usually satisfactory. However, the asymptotical likelihood theory can become inaccurate in small samples. The estimated model may lie outside the stationarity and/or invertibility region and convergence may depend on the initial conditions [6]. Especially for small numbers of observations, action must be taken to improve convergence and to ensure stationarity and invertibility. The same problems occur if poles and zeros of the ARMA process are close to each other.

It remains attractive to use ARMA models because they can be the most parsimonious description of the data. Each estimated parameter gives an increase of 1/N in the mean squared error of prediction. Hence, the adequate representation with the fewest parameters is the best. Many linearised ARMA methods use the parameters of the *first* method of Durbin [7] as initial estimates. A long AR model is used then to reconstruct the innovations or residuals and to use them as regressors in an ordinary least squares procedure. More stages can be added to get a better efficiency by applying the principles of generalized least squares [5,8].

In looking for a safe, robust and practical solution for the ARMA estimation problem, Durbin's *second* iterative method [7] is still another choice, which is computationally simple. This method uses the estimated parameters of a long autoregressive model as the input for a linear MA estimation procedure and computes separately consecutive updates for the MA and for the AR parts. Models estimated with Durbin's second method are always stable and invertible if proper AR and MA algorithms are chosen. This is an important advantage, especially for finite and small samples and for small distances between poles and zeros. Nevertheless, this old and simple ARMA method has not been popular, probably because its accuracy is suspected to be comparable to extended Yule-Walker methods. However, the practical achievable accuracy of Durbin's method is unknown so far, because no theory exists for the optimal choice of long AR orders in finite samples. Durbin's methods are based on the asymptotical equivalence of AR(∞) and ARMA(p,q) or MA(q) processes. Simulations with MA processes, however, have shown that the best AR order is finite and depends on the true process parameters and on the number of observations that is available [4]. Hence, it may be expected that also the accuracy of Durbin's ARMA method can be improved by using a better order for the long intermediate AR model that is used.

In this paper, the theory for best orders in linear regression [9] will be applied to the best AR orders for ARMA processes. Different applications of AR models are: using an AR model for prediction or using the parameters of a long AR model to estimate the MA part of the model. The requirements for the different optimal AR orders will be described theoretically for *known* ARMA processes. The theory about different orders will be verified in simulations. The optimal accuracy of ARMA models obtained with Durbin's second method will be established, for the best long AR order. The theory for known ARMA processes is used to develop a Sliding Window algorithm to select good intermediate AR orders for observations of *unknown* processes.

2. DIFFERENT OPTIMAL AR MODEL ORDERS

The subject of order selection is known as subset selection or as the selection of variables in linear regression. Akaike has introduced AIC for AR model order selection and he described the close connection of AIC with order selection criteria in ordinary regression theory [10]. The size of the best estimated subset depends on the intended use of that subset model [9], which among other things can be:

- prediction, a good fit of the model response to new data
- a small mean squared error of the estimated parameters.

The optimal number of regressors in the best *estimated* model is different for both purposes. That number can be related theoretically to the true values of parameters in known processes, but also to estimated parameter values in practice. However, a theoretical development of the optimal orders is easier for known true processes. Only bias by omission of parameters plays a role then and the variance due to estimation is absent. The analysis depends on the influence of the omission of true values of the parameters on the theoretical residual sum of squares [9], that will

be called RSS_{th} in the sequel. The variance σ^2 of the additive noise is used for scaling. The first application, prediction, forms the basis of almost all existing order selection criteria. Regressors have to be included in the best subset model [9] if addition of each arbitrary group of r regressors causes at least a reduction $r\sigma^2$ in the RSS_{th}, based on the true process parameters. However, if the primary concern is accurate estimates of the parameters, the required reduction of RSS_{th} for any group of r regressors is only σ^2 [9], independent of r. This means that also smaller parameter values will be included in the best subset for parameter accuracy, provided they have as a group the same influence on the RSS_{th} as each single regressor must have in the best subset for prediction. The difference between *true* and *estimated* parameters is that each estimated parameter gives an additional average variance reduction – σ^2 to RSS_{th} (so – $r\sigma^2$ for a group of r regressors).

Some notation is required to apply those results to time series models. An ARMA(p,q) process can be written as

$$\mathbf{A(z)x_n = B(z)\varepsilon_n} \tag{1}$$

with $A(z)=1+a_1z^{-1}+\ldots+a_pz^{-p}$, $B(z)=1+b_1z^{-1}+\ldots+b_qz^{-q}$ and $z^{-1}x_n=x_{n-1}$. Suppose that ε_n represents a series of independent, identically distributed stochastic variables, a white noise process that generates the data; σ_x^2 is the variance of the process and σ_ε^2 is the variance of the innovations ε_n. The theory of regression analysis will be applied to long AR models of a *true* ARMA process (1), defined as:

$$\mathbf{C(z)x_n = \varepsilon_n} \tag{2}$$

with parameters given by C(z)=A(z)/B(z) and with the same innovation variance σ_ε^2 as (1). The true AR model C(z) has order ∞. A finite number of parameters can be determined with any desired accuracy by computing the covariance function of the ARMA process [1,4,11] for some finite order and transforming those covariances to AR parameters or reflection coefficients k_i with the Levinson-Durbin algorithm [2]. The theoretical RSS_{th} of all AR models from order AR(1) to AR(∞) for the given ARMA process (1) can be determined for a sample size of N observations as

$$\mathbf{RSS_{th}(m) = N\sigma_x^2 \prod_{i=1}^{m}(1-k_i^2)} \tag{3}$$

with $RSS_{th}(0)=\sigma_x^2$ and

$$\mathbf{RSS_{th}(\infty) = N\sigma_x^2 \prod_{i=1}^{\infty}(1-k_i^2) = N\sigma_\varepsilon^2}\ . \tag{4}$$

The true k_i describe the bias contribution that is the inaccuracy of all finite order AR models (3) in comparison with (4); estimation variance of parameters plays no role here. The asymptotical expression for the residual sum of squares becomes exactly $N\sigma_\varepsilon^2$ for AR(∞). The best AR order for prediction is found, for given N, as the order K with the property that for arbitrary values of r > 0 :

$$\begin{aligned}\mathbf{RSS_{th}(K+r) &> RSS_{th}(K) - r\sigma_\varepsilon^2}\\ \mathbf{RSS_{th}(K-r) &> RSS_{th}(K) + r\sigma_\varepsilon^2}\ .\end{aligned} \tag{5}$$

So the reduction for r orders is greater than $r\sigma_\varepsilon^2$ below K and smaller than $r\sigma_\varepsilon^2$ above K. This order K can be found by applying an order selection criterion to the sequence of theoretical residuals. In estimation, the variance of each parameter will approximately give an additional expected decrease of RSS equal to σ_ε^2. This leads to the factor 2 in selection criteria, with equal weights assigned to bias and to variance contributions [12]. The best AR order for prediction, for a *known* ARMA process and a given N, is found approximately as the order K with minimum of the generalized information criterion GIC:

$$\mathbf{GIC(K,\alpha) = \ln\left(RSS_{th}(K)/N\right) + \alpha K/N}, \tag{6}$$

with penalty α=1, because only bias is present in $RSS_{th}(K)$.

The AR order with the best parameter accuracy is found as the order for which the theoretical $RSS_{th}(M)$, due to bias of incomplete models, is only σ^2 greater than for AR(∞), so the smallest M with

$$\mathbf{RSS_{th}(M) = N\sigma_x^2 \prod_{i=1}^{M}(1-k_i^2) < (N+1)\sigma_\varepsilon^2} \tag{7}$$

for N observations, or the order with theoretical residual variance less than $(1+1/N)\sigma_\varepsilon^2$, which is the same. The best order obviously increases with N, the number of observations. This theory has successfully been applied to MA processes, where a similar order has been derived and simulations have corroborated the theoretical derivation [4]. The order M is important in Durbin's second ARMA method because the *parameters* of the long AR model are used to estimate the MA parameters.

3. SIMULATIONS WITH FIXED AR ORDERS

Simulations with several AR, MA and ARMA processes have shown that the calculated AR order K, as defined above, is the best order for estimated AR models if they are used for prediction. Durbin's second method consists of a number of simple computations to find an ARMA(p',q') model, with arbitrary p' and q'. For N given observations x_n, the model can be written as:

$$\mathbf{\hat{A}(z)x_n = \hat{B}(z)\hat{\varepsilon}_n}\ . \tag{8}$$

Preparation of Durbin's second method for a fixed AR order i:

- estimate AR(i) model from data x_n with the Burg method [2]
- reconstruct residuals $\hat{\hat{\varepsilon}}_n$ with (2), using AR(i) model for C(z)
- obtain an initial estimate for the p' AR parameters in $\hat{A}^{(0)}(z)$ with Durbin's first ARMA method [7], using p' previous values of x_n and q' previous $\hat{\hat{\varepsilon}}_n$ as regressors
- use in this experiment the i parameters of the estimated AR model to approximate the long C(z) of (2), denoted $C^i(z)$.

Compute $\hat{A}^{(j)}(z)$ and $\hat{B}^{(j)}(z)$, using $\hat{A}^{(j-1)}(z)$: j-1 → j iteratively

- make the long division $D^{(j)}(z)=C^i(z)/\hat{A}^{(j-1)}(z)$
- compute $\hat{B}^{(j)}(z)$, from $D^{(j)}(z)$ with Durbin's MA method [3]
- filter x_n with updated MA model: $\hat{B}^{(j)}(z)w_n = x_n$
- estimate $\hat{A}^{(j)}(z)$ with AR algorithm of Burg, using w_n as data
- if desired, iterate by computing $D^{(j+1)}$ and so on.

The accuracy of estimated ARMA models is expressed in the model error ME [11], which is a scaled version of the prediction error PE. PE is defined as the variance of predictions when an estimated model is applied to new data and

$$\mathbf{ME = N(PE/\sigma_\varepsilon^2 - 1)}. \tag{9}$$

For use in simulations, where the true process parameters are known, an efficient expression in the time domain has been derived for ME [11]. ME can also be used for AR or MA models, by taking zeros for the absent MA or AR parameters. For models of the true structure and with orders p'≥p and q'≥q, the asymptotically achievable minimum of ME equals the number of estimated parameters p'+q', independent of the number of observations [13]. By multiplying with N in (9), ME has been

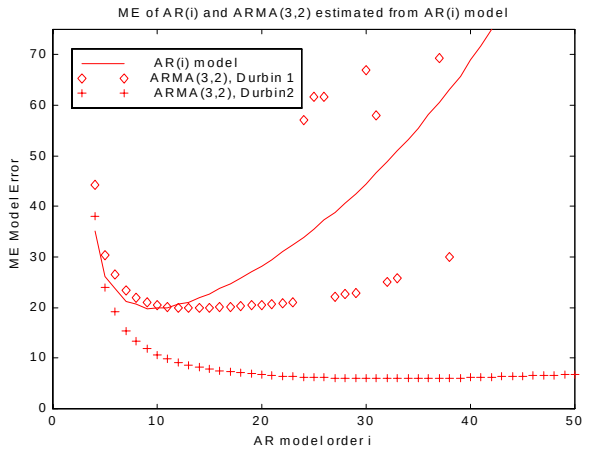

Fig.1 The average Model Error ME of the AR(i) model and two ARMA(3,2) models estimated with Durbin's methods, as a function of the AR model order; 2500 simulation runs, N=100, true ARMA(3,2) process given in (10).

made particularly suited to describe the model quality for different sample sizes.

Fig.1 shows the ME for three different models estimated from the same ARMA(3,2) process:

$$x_n - x_{n-1} + 0.88x_{n-2} - 0.5x_{n-3} = \varepsilon_n + 0.45\varepsilon_{n-1} - 0.5\varepsilon_{n-2} \quad (10)$$

with normally distributed independent innovations ε_n. The average results for AR(i) models and for ARMA(3,2) models estimated from that AR(i) model are presented. The theoretical optimal orders of (5) or (6) for K and (7) for M for prediction and for parameter accuracy are 10 and 31, respectively, for this example and N=100.

The average ME of the AR(10) model is 20.0 and the AR(31) model gives 46.6 if used for prediction. The order 10 turns out to be the best for prediction with an AR model in Fig.1, equal to the theoretical order K. Durbin's first method, applying least squares to delayed observations and reconstructed innovations, shows an irregular average behavior. This is caused by the runs with non-invertible solutions for the MA part. This has to be repaired with constraints on the solution or with mirroring of zeros with respect to the unit circle. Whenever such a reparation is required, the ME values of Durbin's first method become much greater. The best value for ME, 19.9, is found for AR order 12. The problems with non-invertible solutions increase with the AR order i in Fig.1.

Durbin's second method uses the AR parameters of Durbin's first method as initial conditions; no iterations haven been used for Fig.1. The MA part is computed from the parameters of the long AR(i) model. AR order 31, equal to M of (7), gives the minimum ME, 6.0, in Fig.1. Taking lower or higher orders than 31 gives a higher ME, e.g. 10.7 for long AR order 10 and 6.8 for AR(50). It turns out, also in many other examples, that taking lower long AR orders than M can be much worse for the resulting quality of the estimated ARMA model than taking too high orders for the long AR model. The influence of iterations is very moderate. In this example, ME was smallest without iterations.

As far as advises for long AR models orders have been given in the literature, it was the best order for prediction or selected with AIC [10], which is similar. By using that order as the long AR order, the ME of the resulting ARMA model is in most examples much greater than the attainable minimum. By taking the higher theoretical order, the ARMA model with the smallest ME is found.

Fig.1 and many similar simulation results are a verification that the linear regression theory for model orders can be applied to the infinite order AR processes by using (3) for the description of the residual sum of squares as a function of the AR model order. This has been verified for different sample sizes and ARMA(p,q) processes with different p and q. The orders K and M found in simulations agreed always with the theory developed in (5) or (6) and in (7). It is clear that the accuracy of the best ARMA model, expressed in ME (9), is much better than that of the best AR model in this example. The ME value for the best ARMA(3,2) model is 6.0. This is quite close to the value 5 that is the maximum achievable accuracy according to the asymptotical theory [13].

Applying backforecasting [14] in filtering the data to w_n before the AR estimation in Durbin's second ARMA method generally gives lower ME values in examples. So this backforecasting of data before the observation interval is beneficial, as it is in computing the residual variance for an estimated model and it has been used for all results presented. Also in reconstruction of the input signal with a finite approximation of (2) for Durbin's first method, the reconstruction of $\hat{\hat{\varepsilon}}_n$ is better with backforecasting.

4. SLIDING WINDOW ALGORITHM

So far, it has been demonstrated that the two best AR orders are K and M for a *known* ARMA process. This showed that the order M is always greater than (or possibly equal to) K. In practice, the order K, for prediction with the AR model, can be selected with a great variety of order selection criteria. The penalty α in (6) influences the balance between overfit and underfit: taking too high or too low model orders. The smallest value for α, however, without too much statistical probability of overfit, is equal to 2, like in AIC [11]. Hence, that value 2 is used to select a value for K, and not for M. If M would have to be selected directly from data, it would require a penalty α that is almost one. Unfortunately, no order selection criterion can select the desired order M from observations of a process, because the variance of each estimated parameter is equal to the small total bias contribution that has to be detected in (7). A useful order for long AR models can be found in practice with a sliding window technique. This has been applied successfully to long AR models for MA estimation [4]. Details of some possible algorithms are given here. and success depends heavily on details of the implementation, so a precise description is presented.

Sliding Window ARMA(p',q') algorithm: SW

- AR(i) models are estimated for orders 1 until N/2. The order K' is selected with CIC [15] or FSIC [16], order selection criteria that perform well if the highest candidate for selection is greater than N/10. Use of GIC(p,α) of (6) for selection would often erroneously result in selection of an AR order close to N/2, therefore finite sample criteria are preferred.
- Compute initial conditions IC for the p' AR parameters of Durbin's second method with his first ARMA method [7]. O1 is in the tables the intermediate AR order used in Durbin 1.
- If IC fail in practice, because of ill-conditioning of the matrix of regressors or if the AR part of the model is not stationary, the estimated ARMA(p'-1,q'-1) model can be used, denoted IC1
- an improvement has been tested by multiplying IC1 by the ARMA(1,1) model that is found with Durbin's first method from the residuals of the ARMA(p'-1,q'-1), denoted IC2
- The AR order for Durbin's second method is O2, for which a variety of possibilities is tested: SWα is the order αK'+p'+q'

Table 1. The average Model Error and the theoretical optimal AR model orders in simulations for different implementations of Durbin's second ARMA method, as a function of N. O1 and O2 are AR orders used in Durbin's first and second method, respectively.

O1,O2	N=20	50	100	200	N=500	1000	2000	N=5000
K,M	8.39	5.97	5.46	5.49	5.05	5.28	5.10	5.00
M,M	8.91	6.16	5.53	5.37	4.98	5.15	5.07	4.96
K',SW3	8.28	6.14	5.95	6.17	5.59	5.64	5.22	5.04
SW2,SW2	8.28	6.47	6.84	7.13	6.78	6.75	6.08	5.22
SW3,SW3	8.28	5.95	5.82	5.87	5.29	5.35	5.08	4.96
SW3iterated	8.88	7.34	7.50	7.49	6.52	6.79	6.00	5.39
SW4,SW4	8.28	6.07	5.62	5.47	4.86	5.11	4.97	4.99
IC1,SW3	7.18	5.79	6.77	8.38	13.1	22.2	40.0	94.8
IC2,SW3	7.91	5.69	6.12	6.75	7.74	9.83	11.6	14.5
IC2iterated	9.50	7.34	7.57	7.84	7.46	8.28	8.04	7.93
order K	5	7	10	14	21	28	36	48
order M	10	23	31	41	54	64	74	88

- SWα is α times the selected AR order for prediction plus the number of parameters to be estimated. If αK'+p'+q'>N/2, it is replaced by N/2. This choice for the long AR model gives always the possibility to compute the p'+q' ARMA parameters and it is always greater than the *selected* order for prediction K'. This AR order depends on p'+q' of the desired ARMA model. Therefore, the algorithm is called "Sliding Window".
- Backforecasting of the signal before the observation interval is used in filtering w_n to remove the MA part from x_n.
- The influence of iterations has been studied for all variants. It is only reported after 10 iterations for SW3,SW3 and for IC2,SW3 in the tables.

The ME for the optimal theoretical orders K,M or M,M for O1 and O2 have been given in the Tables as a reference to evaluate the effect of *selecting* intermediate AR orders on the ME of the final ARMA model; this is rather small for SW3. Taking different orders, K' and SWα was always worse than taking SWα for both orders, as shown in Table 1 for K',SW3 in comparison with SW3,SW3.

IC2 is an improvement for IC1, but is still worse than the SWα variants. If the initial conditions are given by IC2 and if N is large, iterations are often an improvement in simulations. SW3 with iterations is only better than without iterations in the first columns of Table 2, so iterations are only advisable if IC2 is used for initial conditions; the best number of iterations turns out to depend on N.

Remark that the minimum achievable ME for this example equals 5 in the Cramér-Rao bound of the asymptotical theory. So the Sliding Window algorithms SW3 and SW4 give a very good quality for the estimated ARMA(3,2) models, close to the accuracy that is obtained for the theoretical best intermediate AR order M.

Table 2. The average Model Error for different implementations of Durbin's second ARMA method, N=200, ARMA process with the parameters of (10), except b_1 , as a function of the value for b_1.

b_1 in (10)	-.45	-.30	-.15	0	.15	.30	.45
M,M	10.80	5.78	4.44	4.58	4.42	4.13	5.37
SW2,SW2	10.98	5.68	4.16	4.82	4.55	4.37	7.13
SW3,SW3	10.36	5.81	4.13	4.89	4.79	4.78	5.87
SW3iterated	8.67	5.74	4.37	5.49	5.21	5.61	7.49
SW4,SW4	10.57	5.94	4.22	4.81	4.69	4.90	5.47
IC2,SW3	14.97	7.41	4.02	4.65	5.11	5.99	6.75
IC2iterated	8.88	6.49	4.54	5.52	5.43	6.01	7.84

5. CONCLUDING REMARKS

Theoretical values for two different long AR orders have been derived, that depend on the true ARMA process and on the sample size. The first one is for prediction, the second for parameter accuracy. No selection criterion is available for the second one, but a sliding window algorithm defines a useful compromise for the practice of ARMA estimation with Durbin's second method.

Durbin's second method for estimation of ARMA models is presented with the theoretically optimal AR orders and with practical choices. Durbin's method attained a very good accuracy for ARMA models, close to the lower bound for all sample sizes. It combines computational simplicity and reliability with the large sample accuracy of non-linear maximum likelihood solutions.

6. REFERENCES

[1] M.B. Priestley, '*Spectral Analysis and Time Series'*. London, Ac. Press, 1981.

[2] S.M. Kay and S.L. Marple, 'Spectrum Analysis-A Modern Perspective'. *Proc. IEEE*, vol. 69, pp. 1380-1419, 1981.

[3] J. Durbin, 'Efficient Estimation of Parameters in Moving Average Models'. *Biometrika*, vol. 46, pp. 306-316, 1959.

[4] P.M.T. Broersen, 'The Best Order of Long Autoregressive Models for Moving Average Estimation'. *Signal Processing VIII, Proc. Eusipco Conf., Trieste, Italy*, pp. 799-802, 1996.

[5] B.S. Choi, '*ARMA Model Identification'*, Springer Series in Statistics, New York, Springer-Verlag, 1992.

[6] S. Koreisha and T. Pukkila, 'A Generalized Least-Squares Approach for Estimation of Autoregressive Moving-Average Models', *J. Time Series Anal*., vol. 11, pp. 139-151, 1990.

[7] J. Durbin, 'The Fitting of Time Series Models'. *Revue Inst. Int. de Stat*., vol. 28, pp. 233-243, 1960.

[8] D.Q. Mayne and F. Firoozan, 'Linear identification of ARMA Processes', *Automatica*, vol. 18, pp. 461-466, 1982.

[9] R.R. Hocking, 'The Analysis and Selection of Variables in Linear Regression', *Biometrics*, vol. 32, pp. 1-49, 1976.

[10] H. Akaike, 'A New Look at the Statistical Model Identification'. *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716-723, 1974.

[11] P.M.T. Broersen, 'The Quality of Models for ARMA Processes', *IEEE Trans. on Signal Processing*, vol. 46, pp. 1749-1752, 1998.

[12] P.M.T. Broersen and H.E. Wensink, 'On the Penalty Factor for Autoregressive Order Selection in Finite Samples'. *IEEE Trans. on Signal Processing*, vol. 44, pp. 748-752, 1996.

[13] B. Friedlander and B. Porat, 'A General Lower Bound for Parametric Spectrum Estimation', *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. 32, pp. 728-733, 1984.

[14] G.E.P. Box and G.M. Jenkins, *'Time Series Analysis, Forecasting and Control'*, Revised edition, Holden-Day, San Fransisco, 1976.

[15] P.M.T. Broersen, 'The ABC of Autoregressive Order Selection Criteria'. *Preprints Sysid '97 Conf., Kitakyushu, Japan*, pp. 231-236, July 1997.

[16] Broersen and H.E. Wensink, 'Autoregressive Model Order Selection by a Finite Sample Estimator for the Kullback-Leibler Discrepancy'. *IEEE Trans. on Signal Processing*, vol. 46, pp. 2058-2061, 1998.