

EMPIRICAL ESTIMATION OF PARAMETER DISTRIBUTIONS IN SYSTEM IDENTIFICATION

Wayne J. Dunstan * Robert R. Bitmead *,¹

** Department of Mechanical & Aerospace Engineering,
University of California San Diego,
La Jolla CA 92093-0411, USA.*

Abstract: The distribution of parameter estimates from a finite data record is of concern for assessing the confidence in the resulting estimate. Our interest is in the development of nonlinear dynamical models from experimental data and the problem which arises in associating a degree of confidence with the estimated parameter values. If the distribution of the estimates were known then the variance might provide a sensible measure of confidence. Accordingly we consider here procedures for using a single data record to generate a distribution of parameter estimates.

Keywords: System Identification, Model Confidence, Parameter Distribution.

1. INTRODUCTION

We consider the empirical estimation of the distribution of a parameter estimate from a single data record. This is done by considering several data resampling schemes which are used to generate new data records or subsets of the experimental data from which sample parameter estimates are calculated. The resulting sample distribution of these estimates is used as an approximant of the underlying “true” distribution of the parameter estimate. This true distribution would be revealed in a sequence of independent Monte Carlo trials with repeated calculation of the parameter estimate. However, only the one data set is available and we seek to generate an approximate empirical distribution. The methods which we consider are Sub-sampling, Resampling and Bootstrapping techniques.

The contribution of this paper is in the reformulation of these techniques for empirical estimation of the distribution of the estimated parameters for **dynamical** system models. The consequence of this, is that our models are required to capture the correlation properties of the input-output data as well as possible. We

shall see that this constrains the resampling schemes to preserve the underlying dynamical structure of the data. This results in using the prediction error residuals for the resampling scheme or, in subsampling, the data set. The significance of our study is that it treats these resampling techniques in the context of dynamical modelling.

Finite length data parameter estimates have a level of variability due to the stochastic nature of the data and due to the finite record lengths. Successive trials of the same size would yield different values of the parameter estimate and, if stationary, these values are distributed according to an underlying true distribution. Quantifying the properties of this distribution is an important step in assessing model confidence. A small variance of the parameter estimate indicates high confidence in the model parameters — or at least a high probability that a repeated trial with a similar but independent data set would yield a close parameter value.

This paper proceeds as follows. Section 2 defines the problem of interest. Section 3 describes different methods for empirically estimating the parameter estimate distributions. Section 4 is an example illustrating

¹ Research supported by USA National Science Foundation Grant ECS-0070146 and DARPA Grant N00014-00-1-0799

these methods. Section 5 discusses further extensions of these techniques in System Identification. Section 6 relates the parameter estimate distribution to the overall assessment of model confidence. Some concluding remarks set forth the authors' direction of future research in this area.

2. PROBLEM STATEMENT

A parametric model, $\hat{P}(\theta)$, may be fitted by minimizing a measure of fit, $V(\theta, \mathcal{Z})$, over a parameter set, $\theta \in \Theta$, and a data set, \mathcal{Z} . Formally,

$$\hat{\theta}(\mathcal{Z}) = \arg \min_{\theta \in \Theta} V(\theta, \mathcal{Z}), \quad (1)$$

where $\hat{\theta}(\mathcal{Z})$ is the minimizing parameter estimate and,

$$\mathcal{Z} = (z_1, z_2, \dots, z_N), \quad (2)$$

and, $z_k = \begin{pmatrix} u_k \\ y_k \end{pmatrix}$ where u_k and y_k , $k = 1, 2, \dots, n$ are the input and output vectors of the plant, $P(\theta)$, being identified.

Due to the stochastic nature of experimental data, independent data sets ($\mathcal{Z}^1, \mathcal{Z}^2, \dots$) may yield different minimizing parameter estimates ($\hat{\theta}^1, \hat{\theta}^2, \dots$). These estimates are assumed to originate from the same underlying distribution, $F(\hat{\theta})$. Our aim is to calculate estimates of the statistics of the distribution $F(\hat{\theta})$, such as the variance. For some of the schemes discussed, this will involve computing explicitly an empirical estimate of the distribution function $\hat{F}(\hat{\theta})$.

The approach to producing an $\hat{F}(\hat{\theta})$ is to create a sequence of new data sets $\mathcal{X}^i, i = 1, 2, \dots, M$, from which to compute extra parameter estimates which, under some assumptions, should be distributed as the underlying $\hat{\theta}$. That is, we seek to produce from the experimental data set, \mathcal{Z} , new input-output sets which we combine into each \mathcal{X}^i . Each of these sets should possess similar underlying correlation structure to \mathcal{Z} but should explore a different stochastic variation — thereby generating an estimate of the distribution of identified dynamic model parameters.

3. METHODS FOR PARAMETER DISTRIBUTION ESTIMATION

The approach to produce new sets \mathcal{X}^i will be based on the single data record \mathcal{Z} of length N . To achieve this we shall consider four techniques,

1. Monte Carlo,
2. Subsampling,
3. Resampling, and
4. Bootstrapping.

Because we shall be using a sole \mathcal{Z} , we need to make the following assumptions.

Assumption 1. The data is derived from a stationary process.

Assumption 2. The data are representative of the process variability.

That is, we need to make an ergodic type assumption that the experimental data set is distributed according to the underlying process and that it is sufficiently long to display both the correlation required for system identification and the stochastic variability in the data.

3.1 Monte-Carlo

The Monte-Carlo procedure involves repeating the the experiment M times with independent data sets ($\mathcal{Z}^1, \dots, \mathcal{Z}^M$), which yield different parameter estimates ($\hat{\theta}^1, \dots, \hat{\theta}^M$). These estimates can then be used to create the empirical distribution $F(\hat{\theta})$ or an estimate of its variance.

We introduce Monte Carlo methods here to provide a point of reference for the subsequent schemes. The parameter estimates, $\hat{\theta}^i$, are from the underlying distribution. However, the conduct of such a number of independent trials would require more data than our single set. Accordingly, we consider three different approaches to this problem of generating new data sets based on the single data set \mathcal{Z} .

3.2 Subsampling

From the experimental data set $\mathcal{Z} = (z_1, \dots, z_N)$ we construct M subsampled data sets $\mathcal{X}_S^1, \dots, \mathcal{X}_S^M$ of length W by selecting W -long contiguous subsets of \mathcal{Z} . Thus, if (j_1, j_2, \dots, j_M) are members of the set $(1, 2, \dots, N-W+1)$ then

$$\mathcal{X}_S^i = (z_{j_i}, z_{j_i+1}, \dots, z_{j_i+W-1}). \quad (3)$$

There are $N-W+1$ possible such sequences extractable from \mathcal{Z} .

Since the subsampled sets \mathcal{X}_S^i are taken from the actual data set \mathcal{Z} , they are distributed identically to the original data and possess identical correlation properties. Their length is reduced, however, since $W < N$. Normally one would select $W \sim \mathcal{O}(\sqrt{N})$.

For each \mathcal{X}_S^i a corresponding $\hat{\theta}_S^i$ is identified. The resulting set of parameters ($\hat{\theta}_S^1, \dots, \hat{\theta}_S^M$) are distributed differently from $\hat{\theta}$ because the sample sizes are different. However, subject to assumptions to be clarified shortly, the distributions of $\hat{\theta}_S^i$ and $\hat{\theta}$ are related. Indeed, their means should be the same and their standard deviations scaled by $\frac{1}{\sqrt{W}}$ and $\frac{1}{\sqrt{N}}$ respectively. Provided N and W are large enough, a Central Limit Theorem holds which characterizes the large sample parameter estimate distribution (Ljung, 1999; Soderstrom and Stoica, 1989; Hjalmarsson and Ljung, 1992).

The competing phenomena of the subsampling approach are the limited number $(N-W+1)$ of subsampled trials possible and the requirement for W to be sufficiently long for $\hat{\theta}_S^i$ to exhibit the correct mean and variance behavior. In terms of the data properties, this requires that the correlation and cross-correlation matrices derived from the W -long subsampled sets be close to their stationary values with an error described by the Central Limit Theorem. Thus a sufficiently large W would be determined by the convergence rates for the Law of Large Numbers and the Central Limit Theorem associated with the data. This in turn is linked to strong mixing properties, see (Politis and Romano, 1994).

With subsampling, it is apparent that many subsampled sequences overlap strongly and will yield highly correlated parameter estimates. However, the mixing over the total sample length N together with the order-invariant calculation of the empirical distribution, should see the $\hat{\theta}_S^i$ display the scaled variance properties of the N -sample Monte Carlo distribution. The balance struck between the competing phenomena of sample size, sample number and asymptotic independence is usually to take $W \sim \mathcal{O}(\sqrt{N})$. This is further discussed by Politis (1998).

3.3 Resampling

In resampling we create M new *pseudo-data* sets by randomly sampling from the original N data to form records of length N . Thus,

$$\mathcal{X}_R^i = (z_{i_1}, z_{i_2}, \dots, z_{i_N}), \quad (4)$$

where the indices i_j are selected independently from a uniform distribution of the integers 1 through N . This resampling may be done with replacement or without replacement, in which the latter case corresponds to a permutation set of samples from the original data, referred to as the *Jackknife* (Turkey, 1958; Efron, 1982).

A precept of the resampling method and of the subsequent bootstrapping method is that they require an additional assumption.

Assumption 3. The data set is composed of independent and identically distributed (i.i.d.) samples.

This requirement precludes the possibility of directly applying such a scheme to the sampled input-output data because they are (hopefully) correlated over time. Indeed, our interest is in identifying a dynamical model for the system generating the data. Inclusion of dependent data structures in the resampling and upcoming bootstrap methods normally involves deriving an i.i.d. sequence from the data set (Freedman, 1944; Ljung, 1999; Tjånström, 2000a). For this we use the prediction error residual sequence derived from a high-order model.

Suppose that the data, u_k and y_k , are generated by a linear system with zero mean i.i.d. excitation process e_k . That is

$$y_k = G(z)u_k + H(z)e_k,$$

where $G(z)$ and $H(z)$ are causal, stable linear systems with $H(z)$ stably invertible. Then we may define the associated one-step-ahead prediction of y_k ,

$$\hat{y}_{k|k-1} = \hat{H}(z)^{-1}\hat{G}(z)u_k + (1 - \hat{H}(z)^{-1})y_k$$

and its associated prediction error sequence of residuals,

$$\epsilon_k = \hat{H}(z)^{-1}[y_k - \hat{G}(z)u_k].$$

This residual sequence should be asymptotically white and should yield $\epsilon_k \rightarrow e_k$ as $k \rightarrow \infty$. The sequence ϵ_k will form the basis of the resampling process. We note that the prediction error generation process is simply invertible;

$$y_k = \hat{G}(z)u_k + \hat{H}(z)\epsilon_k.$$

Denote the computed residual set $\mathcal{E} = (\epsilon_1, \dots, \epsilon_N)$, and consider a resampling of this nominally i.i.d. process to yield a new set $\mathcal{E}_R = (\epsilon_{i_1}, \dots, \epsilon_{i_N})$. Here the indices i_j are chosen randomly from a uniform distribution over the integers 1 through N . This then is our resampled residual sequence, whose elements we denote ϵ_k^R .

From \mathcal{E}_R we compute the resampled output sequence,

$$y_k^R = \hat{G}(z)u_k + \hat{H}(z)\epsilon_k^R. \quad (5)$$

Define the pseudo-data element, $z_k^R = \begin{pmatrix} u_k \\ y_k^R \end{pmatrix}$, and construct the pseudo-data set $\mathcal{X}_R = (z_1^R, \dots, z_N^R)$. This data record is then used to produce a parameter estimate $\hat{\theta}_R$.

Repeating this residual resampling process M times yields a sequence of M parameter estimates, which then may be used to compute the mean and variance of $\hat{\theta}_R^i$. Because the residual sequence tends to the true driving noise and this noise is the source of variability of the parameter estimate, this procedure yields another way to generate an approximate distribution of $\hat{\theta}$.

An interesting feature of this approach is that it explicitly relies on knowing a high fidelity model of the underlying process. This, in turn would seem to obviate the need for system identification. However, we shall see that (following the work of Tjånström (2000b)) one may sensibly combine low and high order models to gain an appreciation of model confidence.

3.4 Bootstrapping

Bootstrapping, first proposed by Efron (1979), is another procedure for generating pseudo-data sets from a single data set and is akin to resampling. It also requires an i.i.d. candidate data set. However, in bootstrapping one computes explicitly the distribution of

the residuals ϵ_i , $Q(\epsilon) = \text{Prob}(\epsilon_i \leq \epsilon)$, where as before this vector inequality is taken componentwise.

Rather than resampling the residuals from the original population, bootstrapping uses the empirical distribution function, $Q(\epsilon)$, to generate completely new pseudo-residuals ϵ_k^B with the same distribution. Theoretically, since we are generating completely new i.i.d. pseudo-residual sets, there is no limit to the number of possible new sequences ϵ^B . This should be compared with resampling where there is an absolute (but usually large) limit of N^N possible sequences with replacement and $N!$ sequences without replacement. This is further discussed by Politis (1998) and Zoubir and Boashash (1998).

With bootstrapped sequence ϵ^B we now generate bootstrapped output data,

$$y_k^B = \hat{G}(z)u_k + \hat{H}(z)\epsilon_k^B.$$

Define the pseudo-data element $z_k^B = \begin{pmatrix} u_k \\ y_k^B \end{pmatrix}$, and construct the pseudo-data set $\mathcal{X}_B = (z_1^B, \dots, z_N^B)$. This data record is then used to produce a parameter estimate $\hat{\theta}_B$, which in turn is used to compute the empirical distribution of $\hat{\theta}$.

Similarly to resampling, bootstrapping requires a high-fidelity model of the underlying process. It also necessitates the computation of an explicit estimate of the residual distribution function $Q(\epsilon)$ as an intermediate step.

4. EXAMPLE

We use a System Identification example to compare the following methods,

1. Monte-Carlo,
2. Resampling, and,
3. Subsampling.

Consider a **plant** which is a second order Auto-Regressive (AR) process with measurement noise,

$$\begin{aligned} \begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} &= \alpha \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \\ &\quad (1 - \alpha^2)^{\frac{1}{2}} \begin{bmatrix} \sin(\omega) \\ -\cos(\omega) \end{bmatrix} e_{1,k} \quad (6) \\ y_k &= [1 \ 0] \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + r \cdot e_{2,k} \end{aligned}$$

where $e_{1,k}$ and $e_{2,k}$ are $\mathcal{N}(0, 1)$ sequences of length N . The plant parameters are, $\alpha = 0.8$, $\omega = 0.2$, $r = 1.0$, and $N = 250,000$. A single output data record was generated, (y_1, \dots, y_N) . The first 1000 samples and Power Spectral Density estimate (PSD) of the entire record are shown in Figure (1). The **model** is of a second order AR process,

$$y_k = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}} e_k, \quad (7)$$

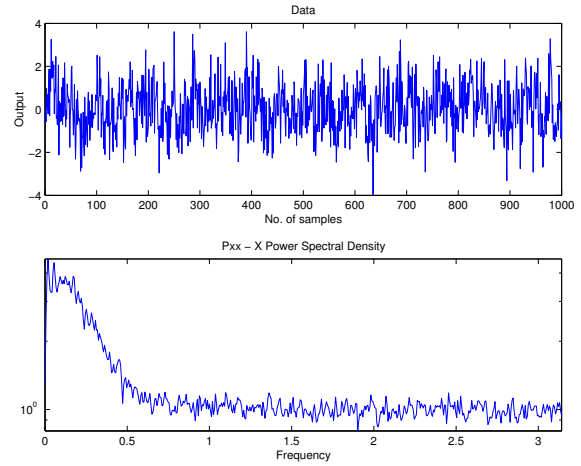


Fig. 1. First 1000 output samples (upper) and PSD of data record (lower).

with $\theta = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$. Our objective is to compute empirical estimates of the variance of $\hat{\theta}$.

4.1 Monte-Carlo

The plant was used to generate $M = 10,000$ independent data sets of length N . A least squares measure was used to identify each $\hat{\theta}^i$. The empirical distribution estimates for $\hat{\theta}_i$ are shown the upper plots in Figure (2) and the variances in Table (1).

4.2 Resampling

The model was fitted to the single plant data set using a least squares estimator. The residual set was resampled 10,000(= M) times. Each resampled residual set of length N was simulated using the parametrized model to produce a resampled output set. New parametrizations were fitted to each resampled output set to identify $\hat{\theta}_R^i$. The empirical distribution estimates for $\hat{\theta}_R^i$ are shown the middle plots in Figure (2) and the variances in Table (1).

4.3 Subsampling

The segment size, W , was set at 500(= \sqrt{N}), which created 249,501(= $N - W + 1$) contiguous segments from the single plant data record. Parametrizations were fitted to each data segment using a least squares estimator. The empirical distribution estimates for $\hat{\theta}_S^i$ are shown the lower plots in Figure (2) with the x -axis scaled for visual comparison with the other distributions. The scaled variances are in Table (1).

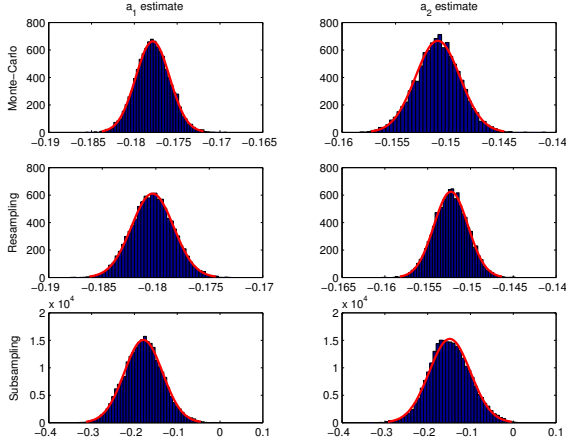


Fig. 2. Empirical distribution estimates for $\hat{\theta}^i$, $\hat{\theta}_R^i$ and $\hat{\theta}_S^i$.

Table 1. Variances of $\hat{\theta}^i$, $\hat{\theta}_R^i$ and $\hat{\theta}_S^i$.

Method	a_1 var	a_2 var
Monte-Carlo	0.0000039	0.0000043
Resampling	0.0000039	0.0000039
Subsampling	0.0000040	0.0000046

Notice that subsampling, requiring no assumptions of i.i.d. data, produced a distribution as close to the Monte Carlo distribution as when resampling was used. This will be paramount to establishing model confidence in systems with no access to white residuals.

5. EXTENSIONS TO SYSTEM IDENTIFICATION

System Identification typically involves producing a low order (L.O.) model that captures the system dynamics present in the data. This objective appears inconsistent with resampling or bootstrapping techniques, which use a high order (H.O.) model to generate i.i.d. pseudo-data. However, we will now outline how L.O. and H.O. models can be used in a *two-stage* fashion to satisfy both objectives.

5.1 L.O./H.O. Method

A two-stage method by Tjånström (2000b) makes connections to Model Error Modelling (Ljung, 1999) by calculating $\hat{F}(\hat{\theta})$ as follows.

The data, (u_k, y_k) , is fitted to a candidate L.O. model which results in non-white residuals,

$$\zeta_k = \hat{H}_1(z)^{-1}(y_k - \hat{G}_1(z)u_k). \quad (8)$$

The set (u_k, ζ_k) is then fitted a H.O. model which results in white (i.i.d.) residuals,

$$\epsilon_k = \hat{H}_2(z)^{-1}(\zeta_k - \hat{G}_2(z)u_k), \quad (9)$$

The residual set $\mathcal{E} = (\epsilon_1, \dots, \epsilon_N)$ is bootstrapped to produce the set $\mathcal{E}_B = (\epsilon_1^B, \dots, \epsilon_N^B)$.

From \mathcal{E}_B the bootstrapped output sequence is computed in two steps,

$$\zeta_k^B = \hat{G}_2(z)u_k + \hat{H}_2(z)\epsilon_k^B \quad (10)$$

$$y_k^B = \hat{G}_1(z)u_k + \hat{H}_1(z)\zeta_k^B. \quad (11)$$

The resulting pseudo-data set $\mathcal{X}_B = (z_1^B, \dots, z_N^B)$ where $z_k^B = \begin{pmatrix} u_k \\ y_k^B \end{pmatrix}$, is used to produce a parameter estimate $\hat{\theta}_B$.

Repeating this residual bootstrapping process M times yields a sequence of M parameter estimates, which then may be used to compute the mean and variance of $\hat{\theta}_B^i$.

Because this method fits the L.O. model first, the mean of $\hat{\theta}_B^i$ will be constrained to the $\hat{\theta}$ of the first L.O. model fit. This however may be of little concern if the goal is estimation of the variance.

5.2 H.O./L.O. Method

This method is similar to the previous one, but does not constrain the mean of $\hat{F}(\hat{\theta})$. It also uses two models, but in the reverse order.

The data, (u_k, y_k) , is fitted to a H.O. model which results in white (i.i.d.) residuals,

$$\epsilon_k = \hat{H}_1(z)^{-1}(y_k - \hat{G}_1(z)u_k), \quad (12)$$

The residual set $\mathcal{E} = (\epsilon_1, \dots, \epsilon_N)$ is bootstrapped (or resampled) to produce the set $\mathcal{E}_B = (\epsilon_1^B, \dots, \epsilon_N^B)$.

From \mathcal{E}_B the bootstrapped output sequence is computed,

$$y_k^B = \hat{G}_1(z)u_k + \hat{H}_1(z)\epsilon_k^B. \quad (13)$$

The resulting pseudo-data set $\mathcal{X}_B = (z_1^B, \dots, z_N^B)$ where $z_k^B = \begin{pmatrix} u_k \\ y_k^B \end{pmatrix}$, is fitted to a L.O. model,

$$y_k^B = \hat{G}_2(z)u_k + \hat{H}_2(z)\epsilon_k, \quad (14)$$

where ϵ_k is an independent white noise sequence. This produces a parameter estimate $\hat{\theta}_B$.

Repeating this “residual bootstrapping and L.O. model fitting” process M times yields a sequence of M parameter estimates, which then may be used to compute the mean and variance of $\hat{\theta}_B^i$.

6. MODEL CONFIDENCE

Our ultimate aim in using these techniques so far discussed is in the assessment of model confidence. Model confidence is related to how uniquely identifiable the best-fit parameter values are from the data and this is captured in part by the distribution of the parameter estimate.

It can be shown under some quite general conditions (P. Stoica and Kay, 1987; Hjalmarsson, 1993) that $\hat{\theta}(V)$ converges to a point $\theta^*(V)$, defined,

$$\theta^*(V) = \lim_{N \rightarrow \infty} \hat{\theta}(V, \mathcal{Z}). \quad (15)$$

Hjalmarsson (1993) shows that

$$\sqrt{N}(\hat{\theta}(V, \mathcal{Z}) - \theta^*(V)) \sim As \mathcal{N}(0, P) \quad (16)$$

where,

$$P = [\bar{V}''(\theta^*)]^{-1} \sum_{\tau=-\infty}^{\infty} R(\tau) [\bar{V}''(\theta^*)]^{-1} \quad (17)$$

where $R(\cdot)$ is the covariance function of the loss function, $\{l'(t, \theta^*)\}$.

The $[\bar{V}''(\theta^*)]^{-1}$ terms in Equation(17) are the inverse sensitivity of the minimizer of V_N at the parameter value $\hat{\theta}$.

It would seem reasonable to want high measure sensitivity to variations in the parameter values, i.e. $\partial^2 V / \partial \theta^2$ large, as this would mean small perturbations in the minimizer estimate adversely affect the measure. This property is considered in Dunstan and Bitmead (2002).

The covariance function, $R(\tau)$, in Equation(17) is the sensitivity of $\hat{\theta}$ to variabilities in the data set, \mathcal{Z} . This variability, due to finite length data and the underlying stochastic processes, we have explored in the distributions of the parameter estimate.

Hence these two properties seem a reasonable beginning in assessment of model confidence in System Identification.

7. CONCLUSION

This paper has formulated some methods for empirical estimation of parameter distributions from finite data records for dynamic models. These methods have been shown applicable in the overall assessment of model confidence.

Future work by the authors' will attempt to apply these results to the practical problem of combustion instability modelling. The model considered for this application is both nonlinear and stochastically driven, and so empirical quantification of the confidence measures seems most to hold the most promise.

REFERENCES

Dunstan, W.J. and R.R. Bitmead (2002). Model confidence fo nonlinear systems. *IFAC World Congress on Automatic Control, Barcelona, Spain*.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7**, 1–26.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *SIAM NSF-CBMS, Monograph* **38**.

Freedman, D. (1944). On bootstrapping two-stage least squares estimates in stationary linear models. *The Annals of Statistics* **12**(3), 827–842.

Hjalmarsson, H. (1993). Aspects on Incomplete Modeling in System Identification. Phd thesis. Linköping University. 409 Department of Electrical Engineering, Linköping, Sweden.

Hjalmarsson, H. and L. Ljung (1992). Estimating model variance in the case of undermodeling. *IEEE Transactions on Automatic Control* **37**(7), 1004–1008.

Ljung, L. (1999). *System Identification: theory for the user*. 2nd ed.. Prentice-Hall, Englewood Cliffs.

P. Stoica, A. Nehorai and S.M. Kay (1987). Statistical analysis of the least squares autoregressive estimator in the presence of noise. *IEEE Transactions on Acoustics, Speech and Signal Processing* **35**(9), 1273–1281.

Politis, D.N. (1998). Computer-intensive methods in statistical analysis. *IEEE Signal Processing Magazine* **15**, 39–55.

Politis, D.N. and J.P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* **22**(4), 2031–2050.

Soderstrom, T. and P. Stoica (1989). *System Identification*. Prentice-Hall International (UK).

Tjänström, F. (2000a). Computing uncertainty regions with simultaneous confidence degree using bootstrap. *IFAC System Identification, Santa Barbara, California, USA*.

Tjänström, F. (2000b). Quality Estimation of Approximate Models. Phd thesis. Linköping University. Department of Electrical Engineering, Linköping, Sweden.

Turkey, J.W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics* **29**, 614.

Zoubir, A.M. and B. Boashash (1998). The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine* **15**, 56–76.