ELSEVIER

Journal of Statistical Planning and Inference 70 (1998) 317–334

journal of statistical planning and inference

Estimating the parameters of the linear compartment model

Debasis Kundu*, Amit Mitra[1]

Department of Mathematics, Indian Institute of Technology, Kanpur 208016, India

Received 14 March 1997; received in revised form 25 November 1997; accepted 26 November 1997

Abstract

In this paper we consider the linear compartment model and consider the estimation procedures of the different parameters. We discuss a method to obtain the initial estimators, which can be used for any iterative procedures to obtain the least-squares estimators. Four different types of confidence intervals have been discussed and they have been compared by computer simulations. We propose different methods to estimate the number of components of the linear compartment model. One data set has been used to see how the different methods work in practice. © 1998 Elsevier Science B.V. All rights reserved.

AMS classification: 62J02; 65D10

Keywords: Compartment model; Confidence intervals; Model selection; Monte Carlo simulations

1. Introduction

Fitting a sum of exponential model to equispaced data, popularly called as linear compartment model, to analyze a real data set is quite common to the scientists and engineers. See, for example, Anderson (1983), Seber and Wild (1989), Gallant (1987), Bates and Watts (1988) and the references therein. In this paper we consider the fitting of the following linear compartment model;

$$y(t)=\sum_{k=1}^{p}\alpha_k e^{\beta_k t}+\varepsilon(t). \tag{1}$$

Here $y(t)$ is observed at the time points $t_1,\ldots,t_n$, which are assumed to be equidistant, α's and β's are unknown parameters and p, is also assumed to be unknown. $\varepsilon(t)$'s are

* Corresponding author. E-mail: Kundu@iitk.ernet.in. Partially supported by a Grant (No. SR/OY/M-06/93) from the Department of Science and Technology, Government of India.

[1] Partially been supported by the National Board of Higher Mathematics, Department of Atomic Energy, Government of India. Presently at the Reserve Bank of India, Mumbai.

PII S0378-3758(97)00190-0

error random variables, they are assumed to be independent and identically distributed (i.i.d.) with zero mean and finite variance σ^2. The problem is, given a sample of size n, $y(t_1),\ldots,y(t_n)$, estimate p, $\alpha_1,\ldots,\alpha_p$ and also estimate $\beta_1,\ldots,\beta_p$. Note that, without loss of generality, we can assume $\beta_1<\beta_2<\cdots<\beta_p$. The ordering may be needed in the identification of the estimates. We also assume $p<\frac{1}{2}n$. This assumption is realistic; in fact, in practice usually n is quite large compared to p.

This is a very old problem and well-known problem in statistical literature. The first reference of this work goes back to the pioneering work of Prony in 1795. After that several articles have appeared mainly to establish some iterative procedures to obtain the least-squares estimators (LSE), when 'p' is known. See, for example, Barham and Drane (1972), Golub and Pereyra (1973), Osborne (1975, 1976), Osborne and Smyth (1986, 1994) and Kahn et al. (1992). It is well known that this problem is a numerically difficult problem. Recently, it is observed by Osborne and Smyth (1994) that the method originally proposed by Osborne (1975) can be very effective, provided the initial value is reasonably good. In fact, it is observed that all the methods are quite sensitive to the initial value. Unfortunately, nowhere in the literature, except the work of Prony (1795), the choice of the initial value has been suggested. In fact, we will see that in many cases the Prony's estimators cannot be used as an initial guess. We discuss a method which can be used to obtain the initial guess of any one of the standard algorithm. We also propose four different types of the confidence intervals and compare their performances through Monte Carlo simulations. We also consider the estimation of 'p'. We propose to use the information theoretic criteria and the cross-validation technique in this purpose. We apply these methods to one real life data set and see how the proposed methods work in practice.

The organisation of the paper is as follows: From Sections 2 to 6, we assume that 'p' is known. In Section 2, we introduce the separable regression technique in this case as it was originally proposed by Richards (1961). We introduce the Prony's estimator in Section 3 and the proposed initial value estimator is presented in Section 4. We discuss different confidence intervals in Section 5 and some experimental results are presented in Section 6. Estimation of 'p' is considered in Section 7 and the data analysis of one data set is being done in Section 8. Finally, we draw conclusions of our work in Section 9.

2. Least-squares estimators

Observe that under the assumption of additive i.i.d. error the LSE can be obtained by minimizing the residual sums of squares,

$$R(\alpha,\beta)=\sum_{t=t_1}^{t_n}\left[y(t)-\sum_{k=1}^{p}\alpha_k e^{\beta_k t}\right]^2, \tag{2}$$

with respect to α and β, where $\alpha=(\alpha_1,\ldots,\alpha_p)$ and $\beta=(\beta_1,\ldots,\beta_p)$. Hence $R(\alpha,\beta)$ can be written in a matrix form as follows:

$$R(\alpha,\beta)=((\boldsymbol{Y}-\boldsymbol{A}(\beta)\alpha)^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{A}(\beta)\alpha), \tag{3}$$

where $\boldsymbol{Y}=(y(t_1),\ldots,y(t_n))$ and $\boldsymbol{A}(\beta)$ is a $n\times p$ matrix of the following form:

$$\boldsymbol{A}(\beta)=\begin{pmatrix} e^{\beta_1 t_1} & \cdots & e^{\beta_p t_1} \\ \vdots & \vdots & \vdots \\ e^{\beta_1 t_n} & \cdots & e^{\beta_p t_n} \end{pmatrix}. \tag{4}$$

Now observe that for a fixed 'p', the LSE of α's are

$$\hat{\alpha}(\beta)=[\boldsymbol{A}^{\mathrm{T}}(\beta)\boldsymbol{A}(\beta)]^{-1}\boldsymbol{A}^{\mathrm{T}}(\beta)\boldsymbol{Y}. \tag{5}$$

Now, if we replace $\hat{\alpha}(\beta)$ obtained from Eq. (5) in Eq. (3), we obtain

$$R(\hat{\alpha},\beta)=Q(\beta)=\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{A}})\boldsymbol{Y}, \tag{6}$$

where $\boldsymbol{P}_{\mathrm{A}}=\boldsymbol{A}(\beta)[\boldsymbol{A}^{\mathrm{T}}(\beta)\boldsymbol{A}(\beta)]^{-1}\boldsymbol{A}^{\mathrm{T}}(\beta)$ is the projection matrix on the space spanned by the columns of $\boldsymbol{A}(\beta)$. Therefore, the LSE of (α,β) obtained by minimizing $R(\alpha,\beta)$ with respect to α,β is same as first obtaining the LSE, $\hat{\beta}$, of β by minimizing $Q(\beta)$ with respect to β and then obtain the LSE of α by putting $\hat{\beta}$ in place of β in Eq. (5). Osborne (1975) considered the minimization of $Q(\beta)$ with respect to β and he transformed the minimization problem to an non-linear eigenvalue problem and proposed an iterative scheme to solve it. Recently it is observed by Osborne and Smyth (1994) that the iterative scheme originally proposed by Osborne (1975) is numerically stable, i.e., if the initial value of the iterative procedure is close to the true value (within the radius of convergence) then the iterative procedure converges almost surely.

3. Prony's estimator

Prony observed the following (see, for example, Barrodale and Olesky, 1981). If

$$\mu_t=\sum_{k=1}^{p}\alpha_k e^{\beta_k t} \quad \text{for } t=t_1,\ldots,t_n, \tag{7}$$

then there exists $(p+1)$ constants, $g_1,\ldots,g_{p+1}$, such that

$$g_1\mu_{t_i}+g_2\mu_{t_{i+1}}+\cdots+g_{p+1}\mu_{t_{i+p}}=0 \quad \text{for } i=1,\ldots,n. \tag{8}$$

Here $\boldsymbol{g}=(g_1,\ldots,g_{p+1})$ is unique upto a constant multiplication and therefore, without loss of generality, we can assume $|\boldsymbol{g}|=\sum_{i=1}^{p+1}g_i^2=1$. Moreover, it can be shown that if $h=t_2-t_1$, then $e^{\beta_1 h},\ldots,e^{\beta_p h}$ are the roots of the following polynomial equation:

$$B(x)=g_1+g_2x+\cdots+g_{p+1}x^{p+1}=0. \tag{9}$$

Therefore there exists a one-to-one correspondence between $\boldsymbol{g}$, such that $|\boldsymbol{g}| = 1$ and $g_1 > 0$, with the nonlinear parameter β. Interestingly, $\boldsymbol{g}$ is independent of α (see Tufts and Kumaresan, 1982). Observe that Eq. (8) can be written in the following form;

$$\boldsymbol{M}\boldsymbol{g} = 0, \tag{10}$$

where $\boldsymbol{M}$ is the $n - p \times p + 1$ matrix of μ's. The rank of the matrix $\boldsymbol{M}$ is p if $n - p > p + 1$ and $\boldsymbol{g}$ is the eigenvector corresponding to the zero eigenvalue of $\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$. This result can be used to obtain an initial estimator of β's from the data. The idea is as follows. Assume that the error variance is small then form a matrix $\boldsymbol{Y}$ from $\boldsymbol{M}$ by replacing μ_t by y_t (assuming $E(y_t) = y_t$) for $t_1, \ldots, t_n$. Then obtain $\hat{\boldsymbol{g}} = (\hat{g}_1, \ldots, \hat{g}_p)$ as the eigenvector corresponding to the minimum eigenvalue of $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$. Normalize $\hat{\boldsymbol{g}}$ s.t. $|\hat{\boldsymbol{g}}| = 1$ and $\hat{g}_1 > 0$. Construct the polynomial equation

$$\hat{B}(x) = \hat{g}_1 + \hat{g}_2 x + \cdots + \hat{g}_{p+1} x^p = 0, \tag{11}$$

and obtain the roots of the polynomial (11) of the form $e^{\hat{\beta}_1 h}, \ldots, e^{\hat{\beta}_p h}$. Note that we can always assume $\hat{\beta}_1 < \hat{\beta}_2 < \cdots < \hat{\beta}_p$, because it is simply renaming the estimators and then the $\hat{\beta}_1, \ldots, \hat{\beta}_p$ can be used as estimators of $\beta_1, \ldots, \beta_p$. It usually behaves quite well as an initial guess provided the error variance is small. If the error variance is large then there are usually two kinds of problems. It may be possible that some roots of the polynomial equation (11) are complex or some roots may be negative. In both the cases it is not possible to obtain the estimators $\hat{\beta}_1, \ldots, \hat{\beta}_p$. Osborne (1975) used these estimators as the initial value for his numerical experiments. In our numerical experiments we observed that Prony's estimator does not exist in some situations.

4. Modified Prony estimator

Prony's idea has been used extensively in the Signal Processing literature to estimate the frequencies of an undamped exponential model. See, for example, the work of Ulrych and Clayton (1976), Kundu (1994a, 1995), Kundu and Mitra (1995a, b), and Tufts and Kumaresan (1982). Several modifications have been suggested by the above authors for the complex-valued signals. Unfortunately, these methods cannot be applied in the case of real data sets. In this section we propose a new estimator which is a modification of the existing Prony's estimator, and it can be used quite effectively as an initial estimator for different algorithms.

First consider the same model as Eq. (1) and y_t and μ_t are defined same as in Section 3. Now choose L, such that, $p < L < n - p$. It can be shown similarly as Tufts and Kumaresan (1982) that there exists $b_0, \ldots, b_L$ such that

$$b_0 \mu_{t_i} + b_1 \mu_{t_{i+1}} + \cdots + b_L \mu_{t_{i+L}} = 0 \quad \text{for } i = 1, \ldots, n - L, \tag{12}$$

where $\sum_{i=1}^{L} b_i^2 = 1$ and $e^{\beta_1 h}, \ldots, e^{\beta_p h}$ are the p roots, among the L roots, of the following polynomial equation:

$$b_0 + b_1 x + \cdots + b_L x^L = 0. \tag{13}$$

Now we can write Eq. (12) in the matrix form as

$$\boldsymbol{A}\boldsymbol{b} = \boldsymbol{0}. \tag{14}$$

Since the rank of the matrix $\boldsymbol{A}$ is p, therefore $\boldsymbol{b}$ is not unique as long as $L > p$ and $L + 1 < n - L$. Observe that Eq. (14) is equivalent to

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b} = \boldsymbol{0}. \tag{15}$$

Consider the spectral decomposition of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ as follows:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \sum_{i=1}^{L+1} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^{\mathrm{T}}, \tag{16}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_{L+1}$ and $\boldsymbol{a}_i$'s are the orthonormal eigenvectors corresponding to λ_i. Since the rank of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is p, this implies $\lambda_{p+1} = \cdots = \lambda_{L+1} = 0$. Therefore, any vector $\boldsymbol{b}$, which belongs to the linear space spanned by $[\boldsymbol{a}_{p+1}, \ldots, \boldsymbol{a}_{L+1}]$ will satisfy Eq. (16). Let us consider the null space of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ as follows:

$$\boldsymbol{N} = [\boldsymbol{a}_{p+1}^{\mathrm{T}} : \ldots : \boldsymbol{a}_{L+1}^{\mathrm{T}}]. \tag{17}$$

Now, since the rank of $\boldsymbol{N}$ is $L - p + 1$, from Eq. (8) it follows that there are $(L - p + 1)$ independent vectors of the following form:

$$\boldsymbol{N}_g = \begin{bmatrix} g_1 & 0 & \cdots & 0 \\ \cdot & g_1 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ g_{p+1} & g_p & \cdots & 0 \\ 0 & g_{p+1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & g_1 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & g_{p+1} \end{bmatrix} \tag{18}$$

such that the space spanned by $\boldsymbol{N}$ and $\boldsymbol{N}_g$ are equal. This idea can be used to estimate $\boldsymbol{g}$ from y_t's as follows.

Replace μ_t's by y_t's to form the matrix $\hat{\boldsymbol{A}}$, and obtain the spectral decomposition of $\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}$ as follows:

$$\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}} = \sum_{i=1}^{L+1} \hat{\lambda}_i \hat{\boldsymbol{a}}_i \hat{\boldsymbol{a}}_i^{\mathrm{T}}, \tag{19}$$

where $\hat{\lambda}_1 > \cdots > \hat{\lambda}_L$ and $\hat{\boldsymbol{a}}_i$ are the orthonormal eigenvectors corresponding to $\hat{\lambda}_i$. Although the rank of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is p, the rank of $\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}$ is $L+1$. Obtain the estimation of the null space of $\boldsymbol{N}$ as

$$\hat{\boldsymbol{N}} = [\hat{\boldsymbol{a}}_{p+1} : \ldots : \hat{\boldsymbol{a}}_{L+1}]. \tag{20}$$

Now we would like to obtain a basis of $\hat{\boldsymbol{N}}$ of the form (18), which can be used to estimate $\boldsymbol{g}$. Let us write

$$\hat{\boldsymbol{N}} = \begin{bmatrix} \hat{\boldsymbol{N}}_1 \\ \hat{\boldsymbol{N}}_2 \end{bmatrix}, \tag{21}$$

where $\hat{\boldsymbol{N}}_1$ is a $p+1 \times L+1-p$ matrix and $\hat{\boldsymbol{N}}_2$ is a $L-p \times L+1-p$ matrix. Since $\hat{\boldsymbol{N}}_2$ is of rank $L-p$, therefore there exists an unique vector $\boldsymbol{b}^1$ upto a constant multiplication such that $\hat{\boldsymbol{N}}_2\boldsymbol{b}^1 = 0$. Therefore, if we denote $\hat{\boldsymbol{g}}^1 = \hat{\boldsymbol{N}}_1\boldsymbol{b}^1$, then

$$\hat{\boldsymbol{N}}\boldsymbol{b}^1 = \begin{bmatrix} \hat{\boldsymbol{g}}^1 \\ \boldsymbol{0} \end{bmatrix}. \tag{22}$$

Proceeding in the same manner, we can say that there exist vectors $\boldsymbol{b}^1, \ldots, \boldsymbol{b}^{L-p+1}$, such that

$$\hat{\boldsymbol{N}}[\boldsymbol{b}^1 : \ldots : \boldsymbol{b}^{L-p+1}] = \begin{bmatrix} \hat{\boldsymbol{g}}^1 & 0 & \cdots & 0 \\ 0 & \hat{\boldsymbol{g}}^2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \hat{\boldsymbol{g}}^{L-p+1} \end{bmatrix}. \tag{23}$$

Now any one or all or a subset of $\hat{\boldsymbol{g}}^k$, for $k = 1, \ldots, L-p+1$, can be used to estimate $\boldsymbol{g}$. It is important to note that when $E(y_i) = y_i$, for $i = 1, \ldots, n$, then $\hat{\boldsymbol{g}}^k = \boldsymbol{g}$, for $k = 1, \ldots, L-p+1$. Now we choose that particular $\hat{\boldsymbol{g}}^k$ from the $L-p+1$, $\hat{\boldsymbol{g}}$'s so that it 'fits' the data best, i.e. for each $\hat{\boldsymbol{g}}^k$; $1 \leqslant k \leqslant L-p+1$, we obtain the estimates of $(\alpha_1, \ldots, \alpha_p)$ and $(\beta_1, \ldots, \beta_p)$ and then calculate the residual sums of squares. We choose that set of α's and β's for which the residual sums of squares is minimum. Although we may get some infeasible roots with respect to some $\hat{\boldsymbol{g}}^k$, but it is observed that not all of them will give infeasible roots.

The method can be easily applied to the following model;

$$y_t = \alpha_0 + \sum_{k=1}^{p} \alpha_k e^{\beta_k t} + \varepsilon_t, \tag{24}$$

that is when the constant term is also present. We apply the above procedure to the data set $y_t - \bar{y}$, when $\bar{y} = \sum_{i=1}^{n} y_{t_i}/n$ if it is known that all the β's are negative. Otherwise, we apply the above method to the complete model, namely

$$y_t = \sum_{k=0}^{p} \alpha_k e^{\beta_k t} + \varepsilon_t \tag{25}$$

and then put $\hat{\beta}_0 = 0$.

Another important aspect is the choice of L. Although in theory L can be any integer satisfying $p < L < n - p$ but it is observed that the performance of the modified Prony estimator depends on L. It is observed (numerically) that as L increases at the beginning the performance of the modified Prony estimator increases in terms of the mean squared errors but after that it starts decreasing. It is observed that for $L \approx \frac{1}{3}n$ the performance is the best, although no theoretical justifications can be provided. It seems that more work is needed in this direction.

5. Confidence intervals

In this section we propose three different confidence intervals other than the classical one as is available in the literature (Gallant, 1987) (p. 105).

One important aspect of this model is its asymptotic properties, namely the consistency and the asymptotic normality. If we take a particular case of this model, namely $t_i = i$, for $i = 1, \ldots, n$ and $p = 1$, $\alpha_1 = 1$, then it follows from Wu (1981) that any estimator of β is inconsistent. In fact to obtain the consistency and the asymptotic normality properties we have to make the boundedness assumptions on t_i's and we have to rewrite the model as that of Kundu (1994b). Under the same set of assumptions as of Kundu (1994b), we have the following result.

If we denote $\theta = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p)$, and $\hat{\theta}$ to be the LSE of θ of the model (1) and θ_0 be the true parameter value of θ in Eq. (1), then we can say:

Theorem 1. *The LSE $\hat{\theta}$ of θ of the model* (1) *is strongly consistent and*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathrm{N}(0, \sigma^2 \boldsymbol{A}^{-1}), \tag{26}$$

where σ^2 is the error variance, $\boldsymbol{A} = ((a_{ij}))$, with

$$a_{ij} = \frac{1}{b-a} \int_a^b h_i'(\theta, t) h_j'(\theta, t)\, \mathrm{d}t, \tag{27}$$

where

$$h_i'(\theta, t) = \frac{\delta}{\delta \theta_i} h(\theta, t) \quad \text{and} \quad h(\theta, t) = \sum_{k=1}^{p} \alpha_k e^{\beta_k t}. \tag{28}$$

Here a and b are the lower and the upper bound of the sampling times. In our case we take $a = t_1$ and $b = t_n$. Now we can use Eq. (26) to obtain $100(1-\alpha)\%$ confidence

interval of θ. Interested readers may refer to Kundu and Mitra (1996) for a detailed proof of the above theorem.

We also propose two Bootstrap confidence intervals. One is the percentile Bootstrap and the other one is the Bootstrap-t confidence interval. Percentile Bootstrap confidence interval can be described as follows:

1. From the sample y_t; $t = t_1, \ldots, t_n$, estimate α's and β's.
2. Estimate $\hat{\varepsilon}_t = y_t - \hat{y}_t$, where

$$\hat{y}_t = \sum_{j=1}^{p} \hat{\alpha}_j e^{\hat{\beta}_j t}$$

   for $t = t_1, \ldots, t_n$.
3. Draw a random sample of size n from $\{\hat{\varepsilon}_{t_1}, \ldots, \hat{\varepsilon}_{t_n}\}$, denote it by $\{\hat{\varepsilon}_{B_1}, \ldots, \hat{\varepsilon}_{B_n}\}$.
4. Obtain the Bootstrap sample $\hat{y}^*_{t_1}, \ldots, \hat{y}^*_{t_n}$, where

$$y^*_{t_i} = \hat{y}_{t_i} + \hat{\varepsilon}_{t_i} \quad \text{for } i = 1, \ldots, n.$$

5. Estimate α's and β's from y^*_t's and denote it by α^*'s and β^*'s.
6. Repeat the process (3)–(5) NBOOT times.
7. From the NBOOT estimators obtain the lower and the upper bound of the $100(1-\alpha)\%$ percentile Bootstrap confidence interval for each α_k and β_k.

We propose the following algorithm for computing the Bootstrap-t confidence intervals.

1. From the sample y_t; $t = t_1, \ldots, t_n$, estimate α's and β's.
2. Estimate $\hat{\varepsilon}_t = y_t - \hat{y}_t$, where

$$\hat{y}_t = \sum_{j=1}^{p} \hat{\alpha}_j e^{\hat{\beta}_j t}$$

   for $t = t_1, \ldots, t_n$.
3. Draw a random sample of size n from $\{\hat{\varepsilon}_{t_1}, \ldots, \hat{\varepsilon}_{t_n}\}$, denote it by $\{\hat{\varepsilon}_{B_1}, \ldots, \hat{\varepsilon}_{B_n}\}$.
4. Estimate α's and β's from y^*_t and also estimate σ^2 as $\hat{\sigma}^2_B$ by mean residual sums of squares.
5. For each α or β (say θ), compute

$$T = \frac{\sqrt{n}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}_B}.$$

6. Repeat steps (3)–(5) NBOOT times.
7. From the NBOOT estimators, obtain the lower and upper bounds of $100(1-\alpha)\%$ Bootstrap-t confidence intervals for each α_k and β_k.

In the next section we have performed some numerical experiments to see how the different methods behave for finite sample.

6. Numerical experiments and discussions

In this section we perform some numerical experiments mainly to see how the modified Prony's estimator behaves compared to the usual Prony estimator as an initial guess for different iterative procedures. We compare Osborne's method, which is well known to work very good for this problem, with that of the usual Gauss–Newton method. We also compare different confidence intervals with respect to their coverage probability and average length. We consider the following model:

$$y_t = -6.0\,e^{-0.232t} + 3.0\,e^{0.0119t} + \varepsilon_t, \quad t = 1,\ldots,n. \tag{29}$$

ε_t is taken to be i.i.d., normal random variables with mean zero and finite variance. We take the values of n as 25, 50, 75 and 100 and the different values of σ are 0.01, 0.05 and 0.1. For each n and σ, we generate 500 samples and compute the Prony estimators (PE) and the modified Prony estimators (MPE). Then using the PE and MPE as the initial guesses we estimate α's and β's using Gauss–Newton (GN) and the Osborne's (1975) (OS) method. We compute the mean squared errors (MSE) of the different estimators. We also report the median number of iteration counts and the maximum number of iteration taken over 500 replications. We also report the number of times the PE does not give the feasible estimators.

We also compute the confidence intervals of the different parameters by four different methods, namely the classical one, the proposed one and the two Bootstrap confidence intervals. We compute the average length and the coverage probability. We take the nominal level to be 90% in our experiments. We report all the results in Tables 1–9. In Tables 1–4, we report the average estimates and the MSEs of the PE, MPE and the LSE. In each box, the first row, the second row and the third row indicate the results of the PE, MPE and LSE, respectively. In each row the first figure represents the average value of the corresponding estimator and in the paranthesis we present the corresponding MSE. Table 5 represents the comparison of the OS method with that

Table 1
Average estimates and the mean squared errors of the different estimators, $N = 25$

σ	α_1	α_2	β_1	β_2
	−6.008(0.0071)	2.999(0.0087)	−0.2323(6.96E − 5)	0.0119(2.16E − 6)
0.01	−6.005(0.0037)	2.999(0.0064)	−0.2321(5.04E − 5)	0.0119(1.53E − 6)
	−6.000(0.0002)	3.000(0.0003)	−0.2320(2.62E − 6)	0.0119(7.34E − 8)
	−6.172(0.2921)	3.135(0.3520)	−0.2354(1.83E − 3)	0.0114(6.22E − 5)
0.05	−6.120(0.1308)	3.044(0.1980)	−0.2349(1.34E − 3)	0.0117(3.83E − 5)
	−6.005(0.0065)	3.002(0.0083)	−0.2324(6.49E − 5)	0.0119(1.86E − 6)
	−6.854(1.321)	3.297(2.321)	−0.2821(1.23E − 2)	0.0120(6.78E − 4)
0.1[a]	−6.321(0.9316)	3.147(1.231)	−0.2524(5.37E − 3)	0.0121(1.40E − 4)
	−6.020(0.0290)	3.012(0.0366)	−0.2329(2.59E − 4)	0.0118(7.69E − 6)

[a] The ordinary Prony estimator did not exist 25 times, so the results of the Prony estimators are based on the average of 475 replications.

Table 2
Average estimates and the mean squared errors of the different estimators, $N = 50$

σ	α_1	α_2	β_1	β_2
	−6.006(0.0070)	3.002(0.0025)	−0.2322(5.08E − 5)	0.0119(2.19E − 7)
0.01	−6.000(0.0030)	2.999(0.0006)	−0.2321(2.57E − 5)	0.0119(4.79E − 8)
	−6.000(1.49E − 4)	3.000(2.44E − 5)	−0.2320(8.97E − 7)	0.0119(1.96E − 9)
	−6.118(0.2038)	3.069(0.0695)	−0.2354(1.31E − 3)	0.0117(5.78E − 6)
0.05	−6.034(0.0637)	2.977(0.0182)	−0.2383(7.14E − 4)	0.0121(1.47E − 6)
	−6.001(0.0038)	3.000(6.07E − 4)	−0.2322(2.26E − 5)	0.0119(4.91E − 8)
	−6.304(1.179)	3.236(0.7069)	−0.2365(5.62E − 3)	0.0109(3.34E − 5)
0.1[a]	−6.168(0.3166)	2.970(0.0671)	−0.2510(3.05E − 3)	0.0122(5.14E − 6)
	−6.003(0.0151)	3.000(0.0024)	−0.2325(9.16E − 5)	0.0119(1.96E − 7)

[a] The ordinary Prony estimator did not exist 11 times, so the results of the Prony estimators are based on the average of 489 replications.

Table 3
Average estimates and the mean squared errors of the different estimators, $N = 75$

σ	α_1	α_2	β_1	β_2
	−5.998(7.18E − 3)	2.998(1.91E − 3)	−0.2323(4.80E − 5)	0.0119(7.82E − 8)
0.01	−6.003(2.53E − 3)	2.999(2.68E − 4)	−0.2324(2.30E − 5)	0.0119(1.06E − 8)
	−6.000(1.49E − 4)	3.000(6.47E − 6)	−0.2320(4.91E − 7)	0.0119(2.38E − 10)
	−6.071(0.2001)	3.041(0.0532)	−0.2356(1.25E − 3)	0.0119(2.10E − 6)
0.05	−6.055(0.0571)	2.991(6.55E − 3)	−0.2373(5.53E − 4)	0.0119(2.60E − 7)
	−6.000(0.0037)	3.000(1.60E − 4)	−0.2322(1.24E − 5)	0.0119(5.92E − 9)
	−6.157(1.127)	3.125(0.3311)	−0.2370(5.45E − 3)	0.0117(1.10E − 5)
0.1	−6.244(0.2670)	2.986(0.0241)	−0.2521(2.42E − 3)	0.0120(9.55E − 7)
	−6.002(0.0151)	3.000(6.34E − 4)	−0.2324(5.03E − 5)	0.0119(2.35E − 8)

Table 4
Average estimates and the mean squared errors of the different estimators, $N = 100$

σ	α_1	α_2	β_1	β_2
	−5.998(7.07E − 3)	2.998(1.15E − 3)	−0.2323(4.63E − 5)	0.0119(2.53E − 8)
0.01	−6.001(2.11E − 3)	3.000(1.40E − 4)	−0.2326(2.23E − 5)	0.0119(2.94E − 9)
	−6.000(1.49E − 4)	3.000(3.77E − 6)	−0.2320(4.43E − 7)	0.0119(7.12E − 11)
	−6.059(0.1959)	3.030(0.0305)	−0.2360(1.21E − 3)	0.0120(6.61E − 7)
0.05	−6.102(0.0411)	3.004(3.72E − 3)	−0.2389(5.48E − 4)	0.0118(7.94E − 8)
	−6.000(0.0033)	3.000(9.41E − 5)	−0.2322(1.12E − 5)	0.0119(1.77E − 9)
	−6.103(1.115)	3.074(0.2213)	−0.2386(5.21E − 3)	0.0119(3.68E − 6)
0.1	−6.349(0.2121)	3.004(0.0129)	−0.2577(1.96E − 3)	0.0119(2.70E − 7)
	−6.002(0.0152)	3.000(3.75E − 4)	−0.2324(4.53E − 5)	0.0119(7.08E − 9)

of the GN method when the starting value is different. We also present the median number of iteration count and the maximum iteration taken for both OS method and GN method, when the initial guess is the usual PE and also when the initial guess is the MPE. In each box, the first figure represents the median iteration count and in

Table 5
Comparison of the Osborne's method and the Gauss-Newton method

N	σ	Prony	Modified Prony
25	0.01	5(8), 5(7)	2(3), 2(3)
	0.05	7(9), 6(9)	3(9), 3(9)
	0.1	15(100), 12(88)	4(41), 4(36)
50	0.01	4(8), 4(6)	2(2), 2(2)
	0.05	6(9), 5(9)	3(3), 3(3)
	0.1	12(70), 10(70)	3(5), 3(4)
75	0.01	4(6), 4(6)	2(2), 2(2)
	0.05	4(9), 4(9)	2(3), 2(3)
	0.1	10(63), 8(63)	3(4), 3(4)
100	0.01	3(3), 3(3)	2(2), 2(2)
	0.05	4(5), 4(5)	2(3), 2(3)
	0.1	7(19), 6(15)	3(4), 3(4)

Table 6
Different confidence intervals, $N = 25$

σ	Methods	α_1	α_2	β_1	β_2
0.01	Asymptotic	0.0527(0.87)	0.0600(0.92)	0.0057(0.93)	0.0009(0.92)
	Classical	0.0494(0.83)	0.0564(0.91)	0.0051(0.90)	0.0009(0.91)
	Bootstrap	0.0497(0.84)	0.0563(0.92)	0.0051(0.91)	0.0009(0.89)
	Bootstrap-t	0.0569(0.91)	0.0648(0.92)	0.0058(0.93)	0.0009(0.93)
0.05	Asymptotic	0.2661(0.87)	0.3012(0.91)	0.0285(0.94)	0.0005(0.91)
	Classical	0.2489(0.82)	0.2827(0.91)	0.0256(0.90)	0.0004(0.91)
	Bootstrap	0.2520(0.85)	0.2845(0.91)	0.0253(0.91)	0.0004(0.89)
	Bootstrap-t	0.2895(0.91)	0.3275(0.92)	0.0290(0.93)	0.0005(0.93)
0.1	Asymptotic	0.5476(0.89)	0.6133(0.92)	0.0571(0.93)	0.0092(0.91)
	Classical	0.5106(0.83)	0.5754(0.91)	0.0512(0.90)	0.0086(0.91)
	Bootstrap	0.5462(0.85)	0.6234(0.92)	0.0527(0.93)	0.0091(0.92)
	Bootstrap-t	0.6290(0.90)	0.7124(0.92)	0.0604(0.92)	0.0104(0.92)

paranthesis the figure indicates the maximum number of iterations required over 500 replications for GN method. Similarly, the second figure represents the result for OS method. In Tables 6–9, we present the comparison of the performances of the four different confidence intervals. In each table we present the average length over 500 replications and in paranthesis we present the coverage probability.

From Tables 1–4, it is observed that as sample size increases or the error variance decreases the MSE of all the estimators, namely PE, MPE and LSE, decrease. It is clear that MPE behave better than PE and the LSE behave better than MPE in all the cases in terms of lower MSE. In our simulations it is observed that although PE does not exist in certain cases but MPE always exist. It is observed that PE may not exist if the sample size is small and the error variance is large. If the sample size is large or the error variance is small PE exist. From Table 5, it is observed that as sample size increases or the error variance decreases, the median number of iteration count or the maximum number of iteration count decreases in all the cases. It is also observed

Table 7
Different confidence intervals, $N = 50$

σ	Methods	α_1	α_2	β_1	β_2
	Asymptotic	0.0493(0.92)	0.0148(0.92)	0.0030(0.88)	0.0001(0.88)
0.01	Classical	0.0399(0.90)	0.0147(0.88)	0.0027(0.87)	0.0001(0.88)
	Bootstrap	0.0404(0.91)	0.0148(0.93)	0.0028(0.87)	0.0001(0.87)
	Bootstrap-*t*	0.0429(0.91)	0.0159(0.91)	0.0029(0.92)	0.0001(0.92)
	Asymptotic	0.2419(0.92)	0.0739(0.88)	0.0151(0.88)	0.0007(0.88)
0.05	Classical	0.2001(0.90)	0.0734(0.86)	0.0139(0.86)	0.0007(0.87)
	Bootstrap	0.2023(0.92)	0.0738(0.84)	0.0139(0.81)	0.0007(0.87)
	Bootstrap-*t*	0.2151(0.92)	0.0793(0.91)	0.0148(0.90)	0.0007(0.93)
	Asymptotic	0.4846(0.91)	0.1480(0.87)	0.0302(0.89)	0.0013(0.89)
0.1	Classical	0.4007(0.89)	0.1469(0.86)	0.0278(0.88)	0.0013(0.87)
	Bootstrap	0.4052(0.91)	0.1480(0.87)	0.0278(0.87)	0.0013(0.87)
	Bootstrap-*t*	0.4316(0.91)	0.1590(0.92)	0.0298(0.93)	0.0014(0.91)

Table 8
Different confidence intervals, $N = 75$

σ	Methods	α_1	α_2	β_1	β_2
	Asymptotic	0.0489(0.92)	0.0081(0.88)	0.0025(0.91)	4.94E − 5(0.88)
0.01	Classical	0.0404(0.93)	0.0082(0.88)	0.0023(0.93)	4.95E − 5(0.88)
	Bootstrap	0.0406(0.95)	0.0081(0.88)	0.0024(0.93)	4.91E − 5(0.88)
	Bootstrap-*t*	0.0422(0.91)	0.0085(0.91)	0.0025(0.92)	5.11E − 5(0.91)
	Asymptotic	0.2447(0.92)	0.0408(0.89)	0.0127(0.91)	0.0002(0.89)
0.05	Classical	0.2021(0.92)	0.0411(0.88)	0.0118(0.93)	0.0002(0.85)
	Bootstrap	0.2030(0.92)	0.0405(0.88)	0.0118(0.93)	0.0002(0.88)
	Bootstrap-*t*	0.2114(0.92)	0.0424(0.89)	0.0123(0.93)	0.0003(0.91)
	Asymptotic	0.4900(0.91)	0.0816(0.88)	0.0255(0.92)	0.0005(0.88)
0.1	Classical	0.4047(0.92)	0.0822(0.88)	0.0236(0.93)	0.0005(0.88)
	Bootstrap	0.4069(0.95)	0.0813(0.88)	0.0237(0.93)	0.0005(0.88)
	Bootstrap-*t*	0.4234(0.92)	0.0849(0.92)	0.0248(0.93)	0.0005(0.91)

that the required number of iteration is much less if the MPE are used as the initial guess to obtain the LSE and it is very prominent if the error variance is high. It is also observed that between the OS method and the GN method the OS method takes less number of iterations compared to GN method in all the cases considered. It indicates that it is better to use the OS method rather than the GN method in general. However, if we use the MPE as the initial guesses, it does not make much differences.

Considering the different confidence intervals, it is observed that they behave quite differently in terms of the average confidence length and the coverage probability. It is observed that as sample size increases or the error variance decreases the average length of the confidence intervals decreases in all the four cases. It is also observed that as sample size increases the coverage probability becomes closer to the nominal value. Among the different methods it is observed that the classical method cannot

Table 9
Different confidence intervals, $N = 100$

σ	Methods	α_1	α_2	β_1	β_2
	Asymptotic	0.0491(0.92)	0.0056(0.88)	0.0024(0.92)	2.51E − 5(0.88)
0.01	Classical	0.0406(0.92)	0.0057(0.86)	0.0022(0.89)	2.52E − 5(0.87)
	Bootstrap	0.0411(0.91)	0.0057(0.87)	0.0022(0.89)	2.49E − 5(0.87)
	Bootstrap-t	0.0422(0.91)	0.0058(0.91)	0.0023(0.90)	2.58E − 5(0.90)
	Asymptotic	0.2459(0.92)	0.0281(0.88)	0.0119(0.93)	0.0001(0.88)
0.05	Classical	0.2034(0.92)	0.0284(0.84)	0.0109(0.89)	0.0001(0.87)
	Bootstrap	0.2057(0.91)	0.0283(0.86)	0.0109(0.88)	0.0001(0.87)
	Bootstrap-t	0.2111(0.91)	0.0292(0.91)	0.0114(0.90)	0.0001(0.90)
	Asymptotic	0.4925(0.92)	0.0562(0.89)	0.0239(0.92)	0.0003(0.89)
0.1	Classical	0.4072(0.92)	0.0569(0.88)	0.0220(0.89)	0.0003(0.88)
	Bootstrap	0.4113(0.91)	0.0565(0.88)	0.0219(0.89)	0.0002(0.85)
	Bootstrap-t	0.4225(0.91)	0.0583(0.90)	0.0227(0.91)	0.0003(0.90)

maintain the nominal coverage probability in many situations, whereas the asymptotic method or the Bootstrap procedure behaves better than the classical method in the sense that the coverage probability is higher than that of the classical one. However in certain situations it also cannot maintain the nominal coverage probability and in certain situations the coverage probability is much higher than the nominal level. Bootstrap-t method works better in that sense, as all the cases considered is capable of maintaining the nominal level although the average length is higher in most of the cases. It is also computer intensive. Comparing all these we recommend to use Bootstrap-t, and if we want to avoid heavy computation, we may use the asymptotic confidence interval.

7. Estimation of p

In this section we propose different methods to estimate p. We assume that $p \leqslant M$, where M is some fixed unknown integer. We propose to use Akaike information theoretic criteria (AIC) and Bayes information theoretic criteria in this setup as it was originally proposed by Akaike (1973), Schwartz (1978) and Risannen (1978) in the general model selection setup. See Kundu and Murali (1996) for the comparison of the different methods. Rao (1988) also suggested to use different information theoretic criteria to estimate the number of signals in a similar problem in Signal Processing. See, for example, Kundu (1992) and Mitra and Kundu (1996) in this connection. AIC and BIC take the following form in this case:

$$\mathrm{AIC}(k) = n \ln \hat{R}_k^2 + 2 \times \text{number of free parameters}, \tag{30}$$

$$\mathrm{BIC}(k) = n \ln \hat{R}_k^2 + \tfrac{1}{2} \ln n \times \text{number of free parameters}. \tag{31}$$

AIC (BIC) chooses that k for which Eq. (30) [Eq. (31)] is minimum. Here $\hat{R}_k^2$ is the residual sums of squares when, the model order is k, i.e.

$$\hat{R}_k^2 = \sum_{t=1}^{n} \left(y_t - \sum_{i=1}^{k} \hat{\alpha}_i e^{\hat{\beta}_i t} \right)^2 . \tag{32}$$

We can use the cross-validation technique also to estimate p. We propose the following procedure for the cross-validation.

1. Take $p = 1$.
2. From the data $y_{t_1}, \ldots, y_{t_n}$, delete y_{t_k}; $k = 1, \ldots, n$.
3. Estimate α's and β's from $y_{t_1}, \ldots, y_{t_{k-1}}, y_{t_{k+1}}, \ldots, y_{t_n}$ by using the missing value technique of Kundu and Kundu (1994).
4. Estimate y_{t_k} by $\hat{y}_{t_k}$ and obtain the cross-validatory error

$$\sum_{k=1}^{n} (y_{t_k} - \hat{y}_{t_k})^2.$$

5. Repeat (2)–(4) for different values of $p = 1, 2, \ldots, M$, and choose that value of 'p' for which the cross-validatory error is minimum.

In the next section we use these techniques to determine the order of the model for one real life data set.

8. Data analysis

In this section we consider one real life data set from Osborne (1972) (p. 185). The data are reported in the following form (t, y_t):

(0, 0.844), (10, 0.908), (20, 0.932), (30, 0.936), (40, 0.925), (50, 0.908),
(60, 0.881), (70, 0.850), (80, 0.818), (90, 0.784), (100, 0.751), (110, 0.718),
(120, 0.685), (130, 0.658), (140, 0.628), (150, 0.603), (160, 0.580), (170, 0.558),
(180, 0.538), (190, 0.522), (200, 0.506), (210, 0.490), (220, 0.478), (230, 0.467),
(240, 0.457), (250, 0.448), (260, 0.438), (270, 0.431), (280, 0.424), (290, 0.420),
(300, 0.414), (310, 0.411), (320, 0.406).

Observe that in this case $n = 33$, $h = 10.0$, $a = t_1 = 0.0$ and $b = t_{33} = 320$. From the plot it is clear that the sum of exponentials can be tried to fit the data. First we use different order determination criteria as proposed in Section 7. Different information theoretic criteria following form shown in Table 10.

Therefore, it is observed from Table 10 that AIC and BIC choose the following model:

$$y_t = \alpha_0 + \alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t} + \varepsilon_t. \tag{33}$$

Observe that Osborne (1975) and Osborne and Smyth (1995) also use the same model to fit this data. They did not mention why they have used that model. Osborne used the initial estimates as $\alpha_0 = 0.5$, $\alpha_1 = 1.5$, $\alpha_2 = -1.0$, $\beta_1 = -0.01$, $\beta_2 = -0.02$. He did

Table 10

Model	R_k^2	NOP	AIC	BIC
α_0	1.1529	2	8.6952	8.1918
$\alpha_1 e^{\beta_1 t}$	0.05105	3	−92.1733	−92.9286
$\alpha_0 + \alpha_1 e^{\beta_1 t}$	0.05089	4	−90.2769	−91.2839
$\sum_{k=1}^{2} \alpha_k e^{\beta_k t}$	0.05082	5	−88.3223	−89.5810
$\alpha_0 + \sum_{k=1}^{2} \alpha_k e^{\beta_k t}$	0.553E − 4	6	−311.4900	−313.0008
$\sum_{k=1}^{3} \alpha_k e^{\beta_k t}$	0.550E − 4	7	−309.6699	−311.4320
$\alpha_0 + \sum_{k=1}^{3} \alpha_k e^{\beta_k t}$	0.525E − 4	8	−309.0169	−311.0310

not mention how he had obtained those initial estimates. With those initial estimates he obtained the following least-squares estimates: $\alpha_0 = 0.3754$, $\alpha_1 = 1.9358$, $\alpha_2 = -1.4647$, $\beta_1 = -0.01287$, $\beta_2 = -0.02212$ after 27 iterations. We obtain the modified Prony estimates with $L = 11$, as follows in this case: $\alpha_0 = 0.1993$, $\alpha_1 = 1.3670$, $\alpha_2 = -0.7157$, $\beta_1 = -0.0087$, $\beta_2 = -0.0261$. Using these values as the initial estimates, the usual Gauss–Newton algorithm converges after five iterations to the same value as that of Osborne. The Osborne's method takes four iterations to converge in this case. The original data and the fitted curve are shown in Fig. 1a. The observed and the fitted values are shown in Fig. 1b. We obtain the following residuals:

−2.499E − 03, 4.611E − 03, 1.171E − 03, −8.564E − 04, −2.649E − 03,
7.331E − 05, −2.650E − 04,
−3.521E − 04, 8.159E − 04, 7.785E − 04, 1.487E − 03, 1.209E − 03,
−5.484E − 04, 1.902E − 03,
−6.150E − 04, −1.783E − 04, 2.069E − 04, −4.138E − 04,
−9.608E − 04, 6.685E − 04, 5.900E − 04,
−1.073E − 03, −1.975E − 04, 3.402E − 04, 6.586E − 04,
8.703E − 04, −9.186E − 04, −6.096E − 04,
−1.111E − 03, 6.592E − 04, −2.205E − 04, 1.318E − 03, 3.384E − 04.

We perform run test (Draper and Smith, 1981) (p. 159) and Durbin–Watson test (Draper and Smith, 1981) (p. 162) on the residuals. It is observed that $z = 0.3594$ and $d = 2.0437$ for run test and Durbin–Watson test, respectively. Therefore, both the tests confirm the independent assumptions of the error random variables.

9. Conclusions

In this paper we consider the linear compartment model as is defined in Eq. (1). We propose MPE to obtain the initial estimators, which is a very important problem. It is observed that the proposed initial estimator can be used for any standard algorithm to obtain the LSEs. We observe that although PE may not exist in certain situation but MPE always exist. We compare the usual GN method with that of the OS method

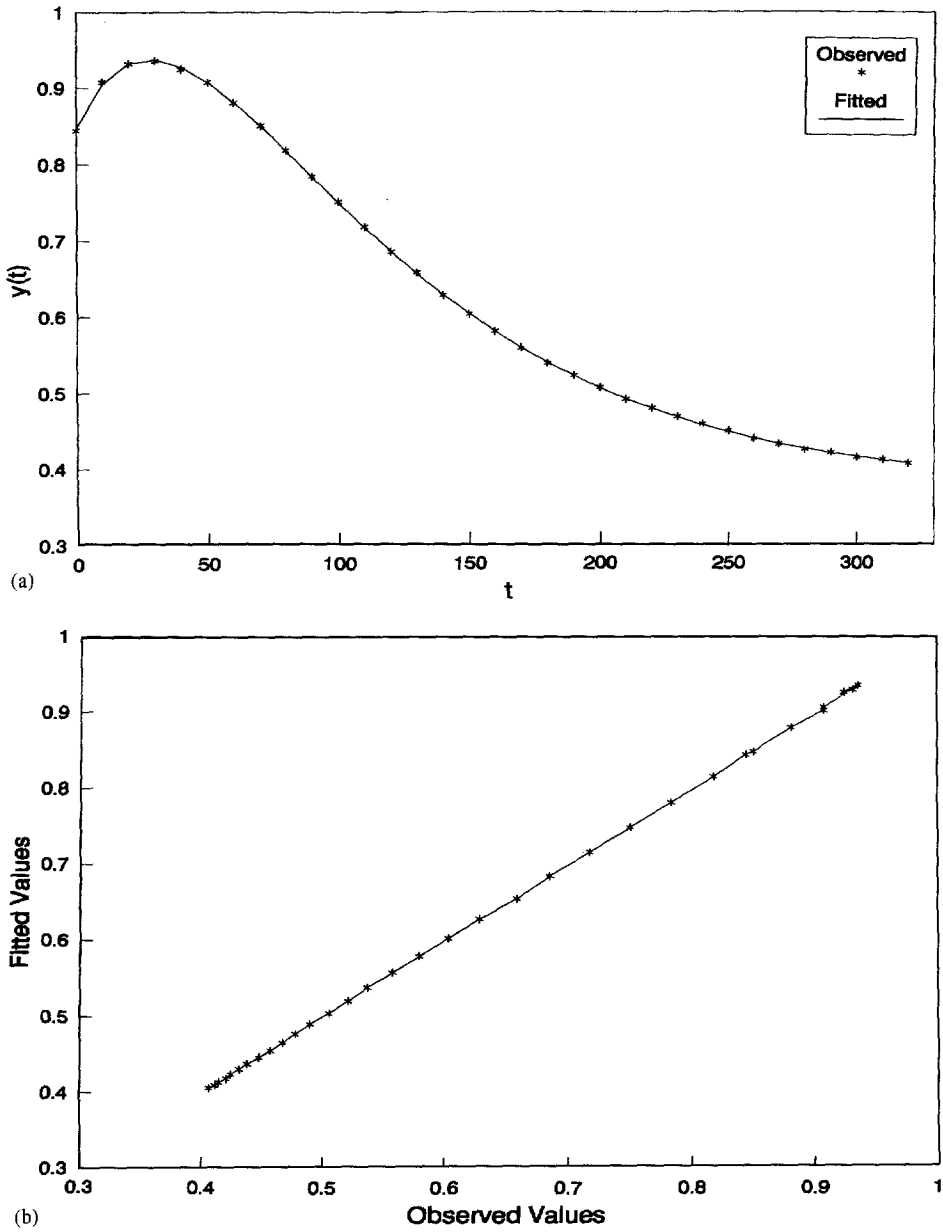

Fig. 1.

and it is observed that the OS method works very well compared to that of GN if the initial guess is PE but if we use the MPE as the initial guesses it does not make much difference, i.e. we do not need any special purpose algorithm to solve this particular problem. We also propose different confidence intervals and it is observed that the

Bootstrap-t works the best in terms of the required coverage probability. We consider the problem of estimating the number of terms in the linear compartment model. We propose to use AIC or BIC to estimate the number of terms. We also provide the scheme to estimate the number of terms by the cross-validation technique. It seems any one of them can be used. Since cross-validation is more computer-intensive, AIC or BIC may be preferred.

Acknowledgements

The authors would like to thank one referee for his constructive suggestions on the earlier draft of this paper.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. Proc. 2nd. Int. Symp. on Information Theory. Academia Kiado, pp. 267–281.

Anderson, D.H., 1983. Compartmental Modelling and Tracer Kinetics. Lecture Notes in Biomathematics. Springer, New York.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley, New York.

Barham, R.H., Drane, W., 1972. An algorithm for least squares estimation of non-linear parameters when some of the parameters are linear. Technometrics 14 (3), 757–766.

Barrodale, I., Oleski, D.D., 1981. Exponential approximation using Prony's method. In: Baker, C.T.H., Phillips, C., The Numerical Solution of Nonlinear Problems, pp. 258–269.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. 2nd ed. Wiley, New York.

Gallant, A.R., 1987. Nonlinear Statistical Models. Wiley, New York.

Golub, G.H., Pereyra, V., 1973. The differentiation of pseudo-inverses and non-linear least squares problems whose variables seperate. SIAM J. Numer. Anal. 10 (2), 413–432.

Kahn, M., Mackisack, M.S., Osborne, M.R., Smyth, G.K., 1992. On the consistency of Prony's method and related algorithms. J. Comput. Graph. Statist. 1 (4), 329–349.

Kundu, D., 1992. Detecting the number of signals for undamped exponential signals using information theoretic criteria. J. Statist. Comput. Simul. 44, 117–131.

Kundu, D., 1994a. Estimating the parameters of complex valued exponential signals. Comput. Statist. Data Anal. 18, 525–534.

Kundu, D., 1994b. A modified Prony algorithm for sum of damped or undamped exponential signals. Sankhya A 56 (3), 525–544.

Kundu, D., 1995. Small sample properties of some parametric and non parametric methods for detection of signals by Monte Carlo simulations. IEE Proc. Vision Image Signal Process. 142 (3), 181–186.

Kundu, D., Kundu, R., 1995. Consistent estimates of superimposed exponential signals when some observations are missing. J. Statist. Plann. Inf. 44, 205–218.

Kundu, D., Mitra, A., 1995a. Consistent method of estimating the superimposed exponential signals. Scand. J. Statist. 22 (1), 73–82.

Kundu, D., Mitra, A., 1995b. Estimating the parameters of Damped exponential signals; A non iterative approach. Signal Process. 46 (3), 363–368.

Kundu, D., Mitra, A., 1996. Fitting a sum of exponentials to equispaced data. Technical Report, I.I.T. Kanpur.

Kundu, D., Murali, G., 1996. Model selection in linear regression. Comput. Statist. Data Anal. 22, 461–469.

Mitra, A., Kundu, D., 1996. Estimating the number of signals for the damped exponential model. Technical Report, I.I.T. Kanpur.

Osborne, M.R., 1972. Some aspects of non-linear least squares calculations. In: Lootsma, F.A. (Ed.), Numerical Methods for Nonlinear Optimization. Academic Press, New York.

Osborne, M.R., 1975. Some special nonlinear least squares problems. SIAM J. Numer. Anal. 12, 571–592.
Osborne, M.R., 1976. Nonlinear least squares, the Levenberg algorithm revisited. J. Austr. Math. Soc. B 19, 343–357.
Osborne, M.R., Smyth, G.K., 1986. An algorithm for exponential fitting revisited. J. Appl. Probab. 418–430.
Osborne, M.R., Smyth, G., 1994. A modified Prony algorithm for fitting sum of exponential functions. SIAM J. Comput. 16 (1), 119–138.
Prony, R., 1795. Essai experimentale et analytique. J. Ecole Polytechnique Paris 1, 24–76.
Rao, C.R., 1988. Some results in signal detection. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics IV, vol. 2. Springer, New York. pp. 319–332.
Richards, F.S.G., 1961. A method of maximum likelihood estimation. J. Roy. Statist. Soc. Ser. B 23, 469–475.
Rissanen, J., 1978. Modelling by shortest data description. Automatica 14, 465–471.
Schwartz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Seber, A., Wild, B., 1989. Nonlinear Regression. Wiley, New York.
Tufts, D.W., Kumaresan, R., 1982. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. IEEE Special issue on Spectral Estimation 70, 975–989.
Ulrych, T.J., Clayton, R.W., 1976. Time series modelling and maximum entropy. Phys. Earth Planetary Interiors 12, 188–200.
Wu, C.F.J., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Math. Statist. 9, 501–513.