A Computer Program for Maximum Likelihood Identification of Linear Multivariable Stochastic Systems

Källstrom, Claes G; Essebo, Tommy; Åström, Karl Johan

1975

Document Version:
Publisher's PDF, also known as Version of record

A COMPUTER PROGRAM FOR MAXIMUM LIKELIHOOD IDENTIFICATION OF LINEAR, MULTIVARIABLE STOCHASTIC SYSTEMS

C G. KÄLLSTRÖM
T. ESSEBO
K J. ÅSTRÖM

Report 7530 (C) December 1975
Department of Automatic Control
Lund Institute of Technology

A COMPUTER PROGRAM FOR MAXIMUM LIKELIHOOD IDENTIFICATION OF LINEAR, MULTIVARIABLE STOCHASTIC SYSTEMS

C G Källström T Essebo K J Åström

Department of Automatic Control
Lund Institute of Technology
Lund, Sweden

1. INTRODUCTION

The maximum likelihood method for parameter estimation has many desirable properties. It is thus a useful tool for systems modelling. Even if the algorithm is motivated using rather strict assumptions like normality of disturbances etc it has a much larger area of applications when estimating parameters in dynamical systems because the method can be interpreted as a modelling of a system in such a way that the prediction error is minimal [1].

This paper describes an algorithm and a computer program for LInear System Parameter IDentification (LISPID), using ML-estimation. The program has been implemented on UNIVAC 1108 and IBM 360. It is flexible and allows for a wide variety of model structures. The model can be given in continuous or discrete time form. Different descriptions of process and measurement noise are permitted. The model can also be time varying. The sampling can be uniform or varying and different types of measurements like instantaneous or integrating are permitted. The model may be parameterized in an arbitrary manner.

The possibility to fit parameters in a continuous time model representation is sometimes advantageous because an application of the basic physical laws gives often a model in this form. In this way it is also easy to handle non-uniform sampling and missing measurements.

Calculation of maximum likelihood estimates is basically an optimization problem. The selection of a proper optimization procedure is an important problem. In this case two algorithms which only use function evaluation were chosen. This is convenient because it allows a great flexibility in the description of the model and its parameterization. It is also possible to use other optimization procedures which require evaluation of gradients and hessians.

There are several clever techniques based on sensitivity derivatives and Lagrange multipliers which can be exploited to compute the derivatives of the likelihood function in special cases [2], [3]. The calculation of derivatives requires, however, analytical computation of the partial derivatives of the difference or differential equations which describe the system. These can be very cumbersome in many cases. For special parameterizations the calculation can, however, be done simply and efficiently leading to efficient numerical algorithms like the one presented in [2]. A discussion of some trade-offs are given in [4]. In the particular algorithm discussed in this paper a flexible model structure is allowed at the expense of increased computations. Because of its flexibility the program is suitable as a general identifier for linear systems. For special applications with a large amount of data it may be more efficient to write a special purpose program.

2. MODEL STRUCTURES

The program LISPID admits many different model structures. The basic model is a process governed by the stochastic differential equation [5]

$$dx = Axdt + Budt + dw \qquad (2.1)$$

where the state vector $x(t)$ has dimension n_x and the control vector $u(t)$ has dimension n_u. The process disturbance $\{w(t),\ t_0 \le t \le \infty\}$ is an n_x-dimensional stochastic process with uncorrelated increments which have zero mean values and

the incremental covariance $R_1 dt$. The initial state of (2.1) is assumed to have mean value m and covariance R_0. It is also assumed that the initial state is uncorrelated with the process noise $\{w(t)\}$.

The program LISPID admits several different models for the measuring process. In all cases it is assumed that the measurements are obtained at discrete times $t_0, t_1, \ldots, t_{N-1}$.

Integrating measurements

The measurement device can be characterized by

$$\begin{cases} dz = Cx\,dt + Du\,dt + de \\ y(t_k) = \int_{t_k}^{t_{k+1}} dz \qquad k = 0,1,\ldots,N-1 \end{cases} \tag{2.2}$$

where $y(t_k)$ is an n_y-dimensional vector of measurements and $\{e(t),\ t_0 \le t \le \infty\}$ is the measurement errors which are assumed to be a stochastic process with uncorrelated increments having zero mean values and incremental covariance $R_2 dt$. The incremental cross covariance between w and e is denoted $R_{12} dt$. A discussion of the integrating measurement procedure is given in [5].

Instantaneous measurements

The program LISPID also admits instantaneous measurements characterized by

$$y(t_k) = \tilde{C}x(t_k) + \tilde{D}u(t_k) + \tilde{e}(t_k) \qquad k = 0,1,\ldots,N-1 \tag{2.3}$$

where the measurement errors $\{\tilde{e}(t_k)\}$ are assumed to be a stochastic process of second order with zero mean and covariance $\tilde{R}_2$. It is furthermore assumed that the measurement errors are independent of w and of the initial state.

Sampling

If the model (2.1),(2.2) or (2.1),(2.3) is sampled and the input signal is constant during the sampling interval, then the following model is obtained (see [5]):

$$\begin{cases} x(t_{k+1}) = \tilde{A}\, x(t_k) + \tilde{B}\, u(t_k) + \tilde{w}(t_k) \\ y(t_k) = \tilde{C}\, x(t_k) + \tilde{D}\, u(t_k) + \tilde{e}(t_k) \\ \qquad\qquad k = 0,1,\ldots,N-1 \end{cases} \tag{2.4}$$

where $\{\tilde{w}(t_k)\}$ is a second order stochastic process with zero mean and covariance $\tilde{R}_1$. The cross covariance between $\tilde{w}$ and $\tilde{e}$ is denoted $\tilde{R}_{12}$. The initial state is uncorrelated with $\tilde{w}$. Algorithms to perform the sampling are described in [6] and [7].

All system and covariance matrices may be time varying in LISPID. It is also possible to estimate parameters of the discrete model (2.4) directly. The program also admits an arbitrary parameterization of the matrices of the models.

3. CRITERIA

The program LISPID admits optimization of different criteria. Assuming that all random variables are normal then the likelihood function L, which is the joint probability density of the outputs $\{y(t_0), y(t_1),\ldots,y(t_{N-1})\}$, is also normal. It is well-known [8] that the logarithm of the likelihood function can be written as

$$-\log L = \frac{1}{2}\sum_{k=0}^{N-1} \log\det R(t_k) + \frac{1}{2}\sum_{k=0}^{N-1} \varepsilon^T(t_k)\, R^{-1}(t_k)\, \varepsilon(t_k) + \frac{1}{2}\, n_y\, N \log 2\pi \tag{3.1}$$

where ε are the innovations

$$\varepsilon(t_k) = y(t_k) - \hat{y}(t_k/t_{k-1}) \tag{3.2}$$

and $\hat{y}(t_k/t_{k-1})$ the one step predictions of the measurements.

The conditional mean $\hat{y}(t_k/t_{k-1})$ and the conditional covariance of residuals $R(t_k)$ can be determined recursively through the Kalman-Bucy filtering theory (see e g [5]):

$$\hat{y}(t_k/t_{k-1}) = \tilde{C}\,\hat{x}(t_k/t_{k-1}) + \tilde{D}\,u(t_k)$$
$$\hat{x}(t_{k+1}/t_k) = \tilde{A}\,\hat{x}(t_k/t_{k-1}) + \tilde{B}\,u(t_k) + K(t_k)\,\varepsilon(t_k)$$
$$\hat{x}(t_0/t_{-1}) = m$$
$$K(t_k) = [\,\tilde{A}\,P(t_k/t_{k-1})\,\tilde{C}^T + \tilde{R}_{12}]\,R^{-1}(t_k)$$
$$P(t_{k+1}/t_k) = \tilde{A}\,P(t_k/t_{k-1})\,\tilde{A}^T + \tilde{R}_1 - \qquad (3.3)$$
$$- K(t_k)[\tilde{A}\,P(t_k/t_{k-1})\,\tilde{C}^T + \tilde{R}_{12}]^T$$
$$P(t_0/t_{-1}) = R_0$$
$$R(t_k) = \tilde{C}\,P(t_k/t_{k-1})\,\tilde{C}^T + \tilde{R}_2$$
$$k = 0,1,\ldots,N-1$$

It is possible to minimize the following simplified loss function instead of (3.1), if the system and covariance matrices are time-invariant and if the sampling rate is constant (see [8]):

$$V_1 = \frac{1}{N} \det \left[\sum_{k=0}^{N-1} \varepsilon(t_k)\,\varepsilon^T(t_k)\right] \qquad (3.4)$$

The parameter values obtained are then the same as if (3.1) is minimized.

The loss function corresponding to (3.1) used in LISPID is normalized:

$$V_2 = -\frac{1}{N} \log L \qquad (3.5)$$

The maximum likelihood estimates of the unknown parameters are obtained, if $\tilde{w}$ and $\tilde{e}$ are sequences of independent normal random vectors and if V_2 is minimum. However, if the gaussian assumptions are not fullfilled, the loss functions V_1 and V_2 may be suitable to use nevertheless.

A modification of the loss function (3.5) can also be used in LISPID to obtain a model that not minimizes the one-step prediction errors only as in the maximum likelihood method, but the p-step prediction errors:

$$V_2(p) = -\frac{1}{N-p+1} \log L(p) \qquad (3.6)$$

where

$$-\log L(p) = \frac{1}{2} \sum_{k=p-1}^{N-1} \log \det S(t_k) + \frac{1}{2} \sum_{k=p-1}^{N-1} \nu^T(t_k) \cdot$$
$$\cdot S^{-1}(t_k)\,\nu(t_k) + \frac{1}{2} n_y (N-p+1) \log 2\pi \qquad (3.7)$$

and

$$\nu(t_{k+p-1}) = y(t_{k+p-1}) - \hat{y}(t_{k+p-1}/t_{k-1})$$

$$\hat{y}(t_{k+p-1}/t_{k-1}) = \tilde{C}\,\hat{x}(t_{k+p-1}/t_{k-1}) + \tilde{N}\,u(t_{k+p-1})$$

$$\hat{x}(t_{i+1}/t_{k-1}) = \tilde{A}\,\hat{x}(t_i/t_{k-1}) + \tilde{B}\,u(t_i), \quad i=k,k+1,\ldots,k+p-2$$

$$P(t_{i+1}/t_{k-1}) = \tilde{A}\,P(t_i/t_{k-1})\,\tilde{A}^T + \tilde{R}_1, \quad i=k,k+1,\ldots,k+p-2$$

$$S(t_{k+p-1}) = \tilde{C}\,P(t_{k+p-1}/t_k)\,\tilde{C}^T + \tilde{R}_2$$

$$k = 0,1,\ldots,N-p \qquad (3.8)$$

$\hat{x}(t_k/t_{k-1})$ and $P(t_k/t_{k-1})$ are obtained from (3.3). Notice that (3.6) is equivalent to (3.5) when p=1. The same modification of the loss function (3.4) becomes

$$V_1(p) = \frac{1}{N-p+1} \det\left[\sum_{k=p-1}^{N-1} \nu(t_k)\,\nu^T(t_k)\right] \qquad (3.9)$$

It is possible to use furthermore generalized loss functions (3.9) and (3.6) by computing an average loss over an interval (p_1,p_2) of prediction steps:

$$V_1(p_1,p_2) = \frac{1}{N-p_2+1} \det\left[\frac{1}{p_2-p_1+1} \sum_{\ell=p_1-1}^{N+p_1-p_2-1} \sum_{k=\ell}^{\ell+p_2-p_1} \nu(t_k)\,\nu^T(t_k)\right] \qquad (3.10)$$

$$V_2(p_1,p_2) = \frac{1}{2(N-p_2+1)(p_2-p_1+1)} \sum_{\ell=p_1-1}^{N+p_1-p_2-1} \sum_{k=\ell}^{\ell+p_2-p_1} \left[\log\det S(t_k) + \nu^T(t_k)S^{-1}(t_k)\nu(t_k)\right] + \frac{1}{2}\,n_y \log 2\pi \qquad (3.11)$$

The previous criteria are obviously special cases of the general loss function (3.10) and (3.11). The ordinary ML-estimate of the parameters of a dynamical system frequently leads to models which are inaccurate in the low frequency region. The generalized loss functions can give models with better low frequency properties.

Due to its flexible structure it is also easy to incorporate an arbitrary user defined loss function into LISPID.

4. LISPID

The program LISPID which is written in FORTRAN consists of 52 subroutines. Including comments the program size is 9 200 statements. The program without any data storage requires a core of 64 k cells on the UNIVAC 1108, if no segmentation is used. Using segmentation and overlays the core required can be reduced to 25 k. Additional memory space is required to store data.

The parameters, which are to be estimated, can enter the system and covariance matrices in an arbitrary way. The mean value m and the covariance R_0 of the initial state $x(t_0)$ can be regarded as parameters to be estimated. A pure time delay of the input signals can also be included and determined. Bias terms of the state and observation equations are easily estimated by introducing an artificial unit step input signal. The dependence of the unknown parameters must be supplied by the user in a special FORTRAN subroutine.

The minimum of the loss function is found by an optimization routine. Since it is extremely tedious to compute the gradient analytically, optimization techniques which only use the values of the loss function have been tried. Two different algorithms are included in LISPID. In the first one the gradient is computed numerically using finite differences. Then a quasi-newton method is applied to find the optimum (see [9]). The other algorithm does not use numerical gradients, but gets information about the loss function by a special search pattern (see [10]).

In the general case with time varying matrices and non-uniform sampling the execution times often become rather long. However, if constant matrices and constant sampling interval are used, the sampling of the continuous time model and the computation of the filter gain K in (3.3) are only needed once for each parameter set. The execution time then decreases. In this case the filter gain K can also be estimated directly. If some measurements are missing of an experiment with uniform sampling rate, LISPID

admits the possibility to save computing time by using the advantages of constant sampling interval and by skipping the contributions from the missing measurements to the loss function.

The execution times are also decreased if the process noise of the model is omitted. The identification technique is then equivalent to an output error method, if the number of prediction steps p is equal to one.

The program LISPID automatically prints and plots the input signals, the measurements, the model outputs and the residuals. The latter ones have in practice proved to be extremely valuable to check measured data. By analysing the autocorrelation functions of residuals and the cross correlation functions between inputs and residuals, which are computed and plotted by LISPID, it can be judged if the model obtained is reasonable.

The appearance of the loss function close to the point proposed as minimum by the optimization algorithm is shown by plotting the loss function, when the estimated parameters are changed one at a time. A possibility of judging if a local minimum point actually has been reached is provided by these plots. Finally, estimated standard deviations of the parameter values obtained can be computed.

5. APPLICATIONS

The program LISPID has been applied to data from many different processes for over a period covering several years. It has also been modified as a result of the experiences gained from these applications. An early version of LISPID was used to determine drum boiler dynamics [11], nuclear power reactor dynamics [12] and the dynamics of a power generator [13]. The program has been extensively used to determine ship steering dynamics. This is reported in [14], [15], [16] and [17]. An application to pharmacokinetics was described in [18] and another biological application is described below. Estimation of macroeconomic models [19] is

another area of application and an excerpt of such an application is also given below.

Plasma kinetics of insulin

Compartment models are frequently used as models for physiological systems. The model

$$\begin{aligned} \frac{dx_1}{dt} &= -k_{21} x_1 + k_{12} x_2 + u \\ \frac{dx_2}{dt} &= k_{21} x_1 - (k_{12} + k_{32}) x_2 \\ y &= x_1 + e \end{aligned} \qquad (5.1)$$

was used to describe the plasma kinetics of insulin in man. See [20]. The experiment was performed by injecting insulin at a constant rate for 2.5 minutes. Samples of the blood were then taken at irregular intervals and the insulin concentration was measured in the samples. An analysis of the measuring procedure also gave the errors of each measurement. The experiment thus gave a sequence of time t_i and concentration measurements $y(t_i)$ with associated estimates of measurement errors.

Since the model (5.1) is one of the standard LISPID models it is straightforward to use the program to estimate the parameters k_{12}, k_{21} and k_{32}. The number of prediction steps was equal to one. In this particular case it was also of particular interest to estimate the parameter

$$k_{el} = \frac{k_{21} \cdot k_{32}}{k_{12} + k_{32}}$$

which represents the total elimination rate. The uncertainty of k_{el} can easily be obtained from the covariance matrix of the original parameters. To see if the data could equally well be explained by a simpler model a first order model ($k_{12} = 0$) was also estimated. The program was applied to a large number of measurements. One example is illustrated in Table I.

The parameter estimates obtained were

$$k_{21} = k_{el} = 0.25 \pm 0.01$$

for the first order model and

$k_{12} = 0.008 \pm 0.008$ $\quad k_{21} = 0.30 \pm 0.05$

$k_{32} = 0.04 \pm 0.07$ $\quad k_{e1} = 0.25 \pm 0.05$

for the second order model. Akaike's criterion for deciding upon a suitable model order was 66.6 and 63 respectively. This indicates that the second order model should be preferred. Notice that the coefficient k_{e1} which is of primary biological interest is comparative well determined even if several of the other parameters are very uncertain. Also notice that estimation of the parameter in a first order model only may give misleading estimates of the uncertainty of the estimate.

Table I. Measured output y, outputs of first and second order models ($\hat{y}_1$,$\hat{y}_2$) residuals (ε_1, ε_2) and standard deviation of measurement errors σ.

t	y	$\hat{y}_1$	$\hat{y}_2$	ε_1	ε_2	σ
4	109	103.0	109.4	6.0	-0.4	3.2
5	77	80.4	82.2	-3.4	-5.2	5.7
6	66	62.8	62.1	3.2	3.9	5.1
7	52	49.0	47.3	3.0	4.7	4.2
8	37	38.3	36.3	-1.3	0.7	3.6
9	27	29.9	28.2	-2.9	-1.2	3.0
10	20	23.3	22.2	-3.3	-2.2	2.7
12	14	14.2	14.3	-0.2	-0.3	2.4
14	9	8.7	9.9	0.3	-0.9	2.1
16	9	5.3	7.3	3.7	1.7	2.1
20	6	2.0	4.8	4.0	1.2	2.9
25	2.5	0.6	3.5	1.9	-1.0	1.9

Econometric model for monetary sector

The program LISPID has been applied to determine macroeconomic models. See [19]. The following model was proposed for the monetary sector

$$M(t) = a_1 M(t-1) + a_2 Y(t) + a_3 B(t) + a_4/P(t) + \tilde{w}_1(t-1)$$

$$(RS-RD)(t) = a_5 (RS-RD)(t-1) + a_6 Y(t) + a_7 M(t) + a_8 + \tilde{w}_2(t-1) \quad (5.2)$$

$$RL(t) = a_9\, RL(t-1) + a_{10}(RS-RD)(t) + a_{11}\, RD(t) + a_{12} + \tilde{w}_3(t-1)$$

where M is real stock of money, Y real GNP, B real monetary base, P GNP-inflator (1968=100), RS and RL short and long term interest rates and RD discount rate. All prices are in mill 1968 Skr.

The vector $\tilde{w} = [\tilde{w}_1\ \tilde{w}_2\ \tilde{w}_3]^T$ consists of process disturbances with covariance matrix

$$\tilde{R}_1 = \begin{pmatrix} |a_{13}| & 0 & 0 \\ 0 & |a_{14}| & 0 \\ 0 & 0 & |a_{15}| \end{pmatrix}$$

The three states M, RS-RD and RL are all measured, and the covariance matrix of measurement errors is

$$\tilde{R}_2 = \begin{pmatrix} |a_{16}| & 0 & 0 \\ 0 & |a_{17}| & 0 \\ 0 & 0 & 0.0001 \end{pmatrix}$$

The variables Y, B, 1/P and RD are inputs. An extra artificial unit step input signal was introduced to estimate the bias terms a_8 and a_{12}.

When rewriting the model (5.2) in standard state space form (cf (2.4)), the parameters above enter the system and covariance matrices in a rather complicated way. This is however easily handled in LISPID. The result of the identification of the model (5.2) to quarterly data about the Swedish economy over the period 1963 to 1972 is shown in Fig 1. The number of prediction steps used in LISPID is equal to one. The parameter estimates obtained are given in Table II.

Table II. Parameter estimates obtained from identification of the model (5.2) to quarterly data about the Swedish economy.

a_1	0.81 ± 0.05	a_{10}	1.0 ± 0.1
a_2	0.00075 ± 0.00113	a_{11}	1.3 ± 0.2
a_3	1.3 ± 0.3	a_{12}	0.53 ± 0.06
a_4	-5100 ± 1200	a_{13}	87 ± 7600
a_5	0.97 ± 0.06	a_{14}	0.0021 ± 0.0043
a_6	0.000016 ± 0.000014	a_{15}	0.0069 ± 0.0130
a_7	-0.00026 ± 0.00044	a_{16}	0.00024 ± 0.64000
a_8	-0.22 ± 0.17	a_{17}	0.0019 ± 0.0035
a_9	0.011 ± 0.100		

6. CONCLUSION

The program LISPID has now been in operation for several years. It has been applied to several parameter estimation problems. Because of the flexibility of the program it has been found to be a very useful tool for estimating parameters in linear dynamical systems.

References

1. Ljung, L: "On consistency for prediction error identification methods". Report 7405 Dept of Automatic Control, Lund Institute of Technology. To appear in Mehra, R K and Lainiotis, D G "Advances in system identification", Marcel Dekker.
2. Åström, K J and Bohlin, T: "Numerical identification of linear dynamic systems from normal operating records". IFAC Symposium on Self-Adaptive Control, Teddington 1965. Also in Hammond, P H (editor) "Theory of self-adaptive control systems", Plenum Press 1966.
3. Gupta, N K and Mehra, R K:"Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations". IEEE Trans AC-16 (1974) 774-783.

4. Åström, K J: "Computational aspects on a class of identification problems". Report Dept EE Univ of Southern California, August 1969.

5. Åström, K J: "Introduction to stochastic control theory". Academic Press, New York, 1970.

6. Mårtensson, K: "Linear quadratic control package, Part II - The discrete problem". Report 6904, Dept of Automatic Control, Lund Institute of Technology, 1969.

7. Källström, C: "Computing exp(A) and $\int \exp(As)ds$". Report 7309, Dept of Automatic Control, Lund Institute of Technology, 1973.

8. Åström, K J and Eykhoff, P: "System identification - A survey". Automatica 7,(1971) 123-162.

9. Fletcher, R: "FORTRAN subroutines for minimization by quasi-Newton methods". Report AERE-R7125, Harwell, 1972.

10. Brent, R P: "Algorithms for minimization without derivatives". Prentice Hall, Englewood Cliffs, New York, 1973.

11. Eklund, K and Gustavsson, I: "Identification of drum boiler dynamics". Proc 3rd IFAC Symp on Identification and System Parameter Estimation, The Hague/Delft, The Netherlands, 1973.

12. Olsson, G: "Modeling and identification of nuclear power reactor dynamics from multivariable experiments". Proc 3rd IFAC Symp on Identification and System Parameter Estimation, The Hague/Delft, The Netherlands, 1973.

13. Lindahl, S and Ljung, L: "Estimation of power generator dynamics from normal operating data". Proc 3rd IFAC Symp on Identification and System Parameter Estimation, The Hague/Delft, The Netherlands, 1973.

14. Åström, K J and Källström, C: "Identification and modelling of ship dynamics". Report 7202, Dept of Automatic Control, Lund Institute of Technology, 1972.

15. Åström, K J and Källström, C G: "Application of system identification techniques to the determination of ship dynamics". Proc 3rd IFAC Symp on Identification and System Parameter Estimation, The Hague/Delft, The Netherlands, 1973.

16. Åström, K J and Källström, C G: "Identification of ship steering dynamics". To appear in Automatica.

17. Åström, K J, Norrbin, N H, Källström, C and Byström, L: "The identification of linear ship steering dynamics using maximum likelihood parameter estimation". Report 7431, Dept of Automatic Control, Lund Institute of Technology, 1974. Also available as Report 1920-1, The Swedish State Shipbuilding Experimental Tank, Gothenburg, Sweden, 1974.

18. Wettrell, G and Andersson, K-E: "Bioavailability and distribution of digoxin in infants". Proc 6th International Congress in Pharmacology, p 127, Helsinki, June 1975.

19. Petersson, E and Wells, C: "Results from single equation and system methods". Arne Ryde Symp on Econometrics, Helsingør, Denmark, 1975.

20. Tranberg, K: Private communication.

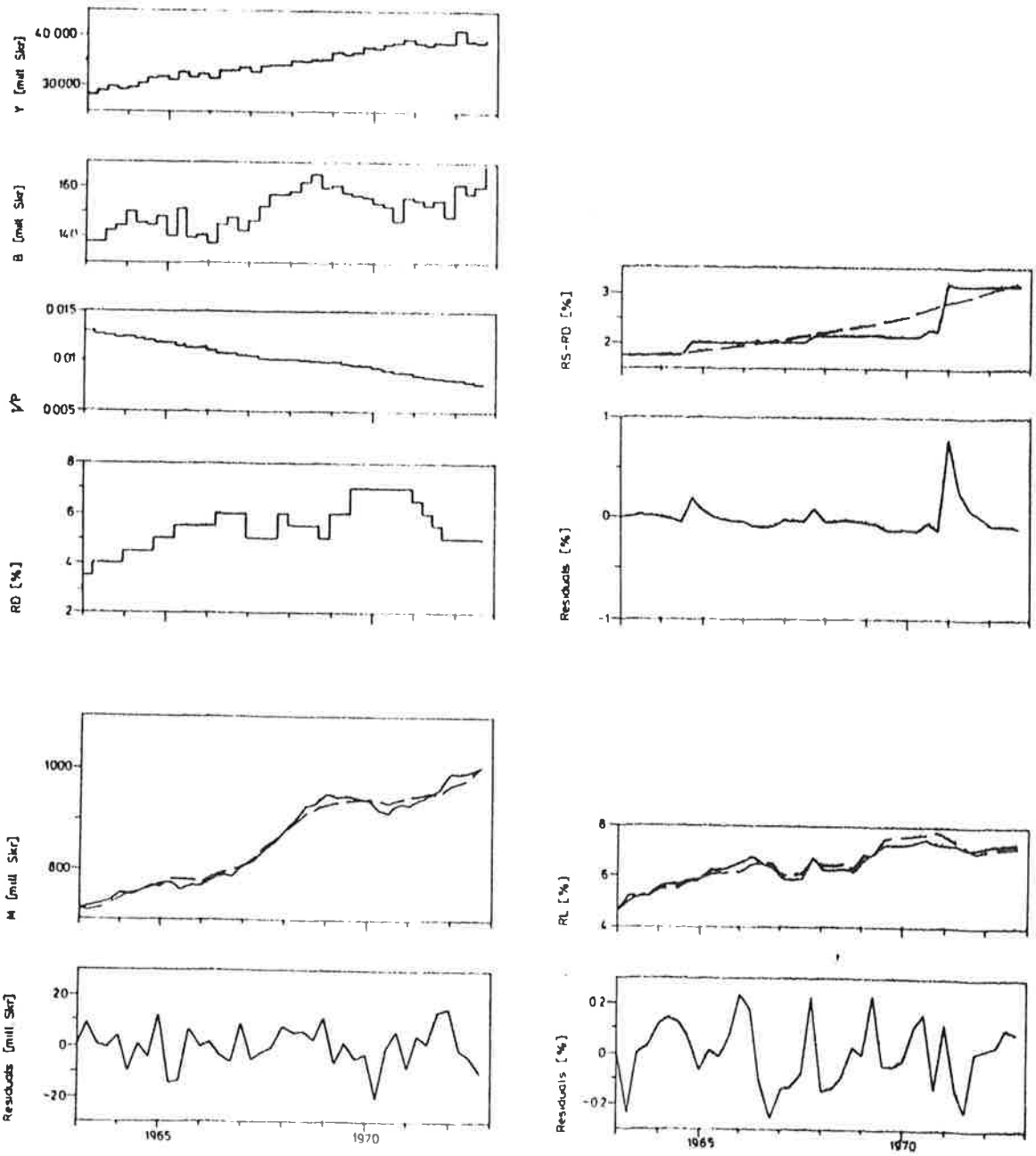

Fig 1. Result of identification of the model (5.2) to quarterly data about the Swedish economy.