System Identification

Paper Presented on
IIASA's 20th Anniversary

Contributors: M. Deistler and I. Vályi

CP-94-5
April 1994

Cover design by Martin Schobel

Printed by Novographic, Vienna, Austria

Contents

Foreword

IIASA celebrated its twentieth anniversary on May 12–13 with its fourth general conference, *IIASA '92: An International Conference on the Challenges to Systems Analysis in the Nineties and Beyond.* The conference focused on the relations between environment and development and on studies that integrate the methods and findings of several disciplines. The role of systems analysis, a method especially suited to taking account of the linkages between phenomena and of the hierarchical organization of the natural and social world, was also assessed, taking account of the implications this has for IIASA's research approach and activities.

This paper is one of six IIASA Collaborative Papers published as part of the report on the conference, an earlier instalment of which was *Science and Sustainability*, published in 1992.

The term "identification" came into use by economists in the late 1920s, but the general idea has existed at least as long as the use of mathematics in science, and is applicable to natural as well as social phenomena. Identification is finding the underlying structure that generated the observed data. In fact we never find the structure itself, but at best a model that is uniquely capable of doing what the structure does.

Usually identification is sought by solving for the values of parameters in a given set of equations – often linear equations. Professor Deistler would broaden the search beyond finding the right coefficients in a set of linear equations; his method permits the use of intuition as well as fitting to find the most likely model.

Committee for IIASA '92
Nathan Keyfitz (Chair)*

*Members of the Committee for IIASA '92 were: Nathan Keyfitz (Chair), Peter E. de Jánosi, Alexander Kurzhanski, Arkadii Maltsev, Nebojša Nakićenović, Roderick Shaw, Claudia Heilig-Staindl, Evelyn Farkas

System Identification

Manfred Deistler

1 General Remarks

The problem of identification is to obtain good models for (real or artificial) phenomena from data. Thus identification is modeling where not only theory but also data are used. The task of identification is often so complex that it cannot be performed in a naive way with the naked eye. In addition many identification problems share common features. For these reasons, methods and theories have been developed which make system identification a subject on its own.

Nevertheless, system identification has many different aspects and facets depending in particular on the kind and amount of *a priori* information the candidate systems used and on the intended use for the identified model. Identification may be performed for the following purposes:

- for encoding data (by system parameters) or for giving a "non-theoretical" description of the relation between data;
- for spectral estimation;
- for prediction, filtering or interpolation;
- for analysis and simulation of systems;
- for control;
- for estimation of parameters in models obtained from "physical" theories;
- for discriminating between conflicting theories.

Institute for Econometrics, Operations Research and Systems Theory, Technical University of Vienna, Vienna, Austria.

System identification has a wide range of applications in many "empirical" branches of science. Important areas of application are:

- Signal processing, in particular speech processing, sonar and radar applications.
- Identification of technical systems (plants) for the purpose of control.
- Modeling of technical systems for simulation in order to avoid, or for the design of, "real" experiments: monitoring of technical systems.
- Economic and business applications, in particular econometrics, e.g., testing of economic theories, estimation of "deep" parameters, forecasting and policy simulation with macromodels, analysis and forecasting for financial data.
- Ecological applications, e.g., modeling the dynamic behavior of ecosystems.
- Geophysical applications, e.g., analysis of seismic signals.
- Biological and medical applications, e.g., analysis of EEG data.

Given this wide range of applications, it is not surprising that approaches to identification have been developed in a number of different and partly widely separated areas such as system and control engineering, signal processing, statistics (in particular time series analysis), econometrics or in certain fields of applied mathematics. It is also not surprising, that there is no unified theory of system identification; nevertheless there is a rather complete theory for identification of linear systems from discrete time series data in the mainstream case. We will describe some of the basic features of this theory in this paper. Then some remarks concerning alternative approaches for linear system identification and identification of nonlinear systems are made.

The mainstream case has a "prototype character" for other cases too, since a number of features of the mainstream case also appear in other cases. In the actual identification in many cases, in particular if careful modeling is required, the procedures are not completely automatized. Formulation of *a priori* information, preprocessing of data, interpretation and evaluation of statistical results and an interactive modeling strategy which combines the statistical tools in an intelligent way require a lot of human interference and also knowledge and experience. Recently, also, artificial intelligence approaches are used, partly in order to replace or to systematize human interference in system identification.

2 Some Remarks Concerning the History of the Subject

Early systematic approaches to time series analysis date back to the end of the 18th and the 19th century. The main focus at that time was on the search for hidden periodicities and trends, e.g., in the orbits of the planets and the moon. The search for such unobserved components occupied great mathematicians such as Euler, Fourier, Lagrange and Laplace. Subsequently, the periodogram as an instrument to search for hidden periodicities was introduced by Stokes and used by Schuster.

Moving average (MA) and autoregressive (AR) systems for time series modeling were introduced by Yule in the 1920s; one main aim was to model "non exact periodicities" such as business cycles.

In the 1930s and 1940s, the linear theory of weak sense stationary processes was developed by Cramer, Kolmogorou, Wiener, Wold and others. Cornerstones were results on spectral representation, Wold representation, spectral factorization and linear least squares prediction and filtering. This theory is based on population second moments, rather than on data, and thus is not part of statistics in the narrow sense; it still constitutes one of the main foundations for time series analysis.

A rather modern approach to system identification was taken in the work of the Cowles Commission (T.W. Anderson, Haavelmo, Klein and Koopmans and others) in early econometrics. Here for the first time a systematic approach to identifiability and maximum likelihood estimation for multi-input multi-output (MIMO) systems was developed. However, this was under the restricting assumptions of white noise equation errors and known dynamic specification. The first macroeconomic models were estimated by Tinbergen and Klein.

After the Second World War, methods for estimating spectra and transfer functions have been developed; this development has been triggered particularly by Tukey.

At about the same time least-squares-type and maximum-likelihood-type estimators for AR and autoregressive moving average (ARMA) systems have been investigated, in particular with regard to their asymptotic properties and mainly for the single-input single-output (SISO) case with given dynamic specification. These results are contained in the now classical books by Anderson (1971) and Hannan (1970).

The introduction of state space models and the subsequently developed structure theory for MIMO systems, triggered by Kalman in the 1960s constituted another important step in system identification.

The book by Box and Jenkins (1970) had a substantial influence on applications. It provided an "integrated" approach, including procedures for detrending the original data and for (non-automatic) determination of lag lengths from data. The Box–Jenkins approach has been mainly developed for the single-output case.

A further important step was the development of (automatic) procedures for estimating the dynamic specification. In particular, the estimators based on information criteria introduced by Akaike and Rissanen and the investigation of their asymptotic properties by Hannan should be mentioned here.

3 The Main Stream Theory for Linear System Identification

The term "mainstream theory" was introduced in Deistler (1989) in order to describe a certain setting for the problem. Both with respect to the existing body of methods and theories and with respect to the range of applications, linear system identification is a rich and extensive subject now. This is also documented by a number of recent books [Caines (1988), Hannan and Deistler (1988), Ljung (1987), Söderström and Stoica (1988), see also the related book by Brockwell and Davies (1991)].

The basic setting for mainstream theory is as follows:

1. The model class, i.e., the set of all *a priori* candidate systems to be fitted to the data, consists of linear, finite dimensional, constant parameter, causal and stable systems only; the classification of the variables into inputs and outputs is known *a priori*.
2. Stochastic models, in particular stationary ergodic processes with rational spectral densities are used for the modeling of noise. Thus we are in the realm of inferential statistics, and we can evaluate estimation and testing procedures.
3. The inputs are assumed to be noise-free. The noise is orthogonal to the inputs and is added to the outputs or to the equations.
4. The criteria for goodness of fit of the system to the data are of the (Gaussian) maximum likelihood (ML) type.
5. In many cases, the model class will be so large that, basically due to problems of overfitting, estimators obtained from optimizing "goodness" of fit only, will be deficient in certain respects. In these cases, the model class is decomposed into (finite dimensional) subclasses. Each subclass is described by its so-called dynamic specification expressed by a vector of integers (a so-called multi-index, e.g., the maximum lag lengths). In general, the dynamic specification has to be determined from data too; here this is done by optimizing a criterion taking into account both the best quality of fit within the subclass and the complexity of the subclass, described by its dimension. Once the subclass is fixed, we have a parametric estimation problem, where the estimates are obtained by optimizing the fit. The overall approach (which includes estimation of the multi-index) could thus be called semi-nonparametric.
6. As far as the desirable properties of the estimators are concerned, the main focus is on consistency and asymptotic efficiency.

4 The Main Stream Case – Structure Theory

There are several different representations for linear systems (see e.g., Hannan and Deistler, 1988). The *input-output representation* is of the form

$$y_t = \sum_{j=0}^{\infty} K_j u_{t-j} \quad ; \quad K_j \in \mathbf{R}^{s \times m} \tag{1}$$

where y_t are the outputs, u_t are the inputs and K_j are the weighting coefficients. Let z denote the complex variable as well as the backward shift on the integers; the *transfer function* is of the form

$$k(z) = \sum_{j=0}^{\infty} K_j z^j \ .$$

For the sake of simplicity here we restrict ourselves to the case where the inputs are (unobserved) s-dimensional, white noise ε_t (i.e., $E\varepsilon_t = 0$, $E\varepsilon_s\varepsilon_t = \delta_{st} \cdot \Sigma$). The *ARMA representation* is of the form

$$a(z)y_t = b(z)\varepsilon_t \tag{2}$$

where

$$a(z) = \sum_{j=0}^{p} A_j z^j \ ; \ A_j \in \mathbf{R}^{s \times s} \ ; \ b(z) = \sum_{j=0}^{q} B_j z^j \ ; \ B_j \in \mathbf{R}^{s \times s} \tag{3}$$

are polynomial matrices. In the main stream case we impose the stability assumption $\det a(z) \neq 0$, for $|z| \leq 1$, and the miniphase condition $\det b(z) \neq 0$ for $|z| < 1$. The matrices A_j, B_j and Σ contain the real-valued parameters for the system and the noise respectively, p and q are examples of integer parameters. The transfer function of (2) is given by $k(z) = a^{-1}(z)b(z)$. The process (y_t) is stationary. Its covariance function $\gamma(s) = Ey_s y_0'$ is in one-to-one relation with the spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma(s) e^{-i\lambda s} = \frac{1}{2\pi} k(e^{-i\lambda}) \cdot \Sigma \cdot k^*(e^{-i\lambda}) \tag{4}$$

where * denotes the conjugate-transpose.

Another important class is state space representations, which have been popularized by the seminal work of Kalman (see, e.g., Kalman, 1963). They are of the form

$$x_{t+1} = Fx_t + G\varepsilon_t \tag{5}$$

$$y_t = Hx_t + \varepsilon_t \tag{6}$$

where x_t is the n-dimensional state, $F \in \mathbf{R}^{n\times n}$, $G \in \mathbf{R}^{n\times s}$, and $H \in \mathbf{R}^{s\times n}$ are parameter matrices (containing the real-valued parameters) and n is an integer parameter. The transfer function is given by

$$k(z) = H(I - Fz)^{-1}zG + I$$

Since ARMA and state space representations are equivalent in a certain sense, we will only consider the ARMA case here. Every linearly regular stationary process can be approximated with arbitrary accuracy by an ARMA process (by an appropriate choice of p and q). This, together with the fact that for given p and q values (or other integers determining the dynamic specification) estimation of (the real-valued) ARMA parameters is a parametric estimation problem, is the main reason for the widespread use of ARMA (and state space) systems in identification. A stationary process is an ARMA process (or equivalently coming from a state space system) if and only if it has a rational spectral density. Thus ARMA identification means rational approximation of spectral densities.

Of particular interest are AR systems, i.e., ARMA systems satisfying $B_j = 0$, $j > 0$, mainly because identification in this case is simpler and numerically faster.

Structure theory is concerned with the relation between external characteristics [here the population second moments of the observations (y_t)] and the internal parameters (here the real and integral parameters of the ARMA system and the noise parameters in Σ). This relation, in our case, is described by equations (3) and (4).

The problem may be decomposed into two steps. The first is to determine the transfer function $k(z)$ and Σ from f. Under the assumptions $k(0) = I$ and $\Sigma > 0$ this can be done in a unique way.

The second problem, which is the main problem, is concerned with the relation between $k(z)$ and $a(z)$, $b(z)$.

For the case where no (additional) *a priori* information is available, the model classes considered are T_A, the set of all pairs $(a(z), b(z))$ satisfying the stability and miniphase conditions, and $a(0) = b(0) = I$ (for fixed s, and arbitrary p, q values) and the corresponding set U_A of transfer functions $k(z)$. Let $\pi : T_A \rightarrow U_A$ denote the mapping defined by $\pi(a, b) = a^{-1} \cdot b$. Then one problem is that π is not injective, and thus in order to ensure identifiability [i.e., uniqueness of the ARMA system (a, b) for a given transfer function k], suitable representatives from the equivalence classes $\pi^{-1}(k)$ have to be selected. The second, even larger problem is that, in general, no continuous selection of representatives for all $k \in U_A$ exists. Thus U_A and T_A are broken into (finite dimensional) subclasses U_α and T_α, say, which allow for a continuous parametrization $\varphi_\alpha : U_\alpha \rightarrow T_\alpha$, attaching to every $k \in U_\alpha$ the unique, corresponding values $(a, b) \in T_\alpha$ [i.e., $\pi(a, b) = k$]. In general, α is a multi-index of integers, in the simplest cases we have $\alpha = (p, q)$ or $\alpha = \max(p, q)$. Continuity of φ_α is important for the "well posedness" of the problem of parameter estimation. In particular, consistency of transfer function estimators then implies consistency for the estimators of (a, b). In general, for $(a, b) \in T_\alpha$ some of the entries in A_j, B_j will satisfy certain restrictions. By $\tau \in \Re^{d\alpha}$ we denote the d_α dimensional, say, vector of free parameters for $(a, b) \in T_\alpha$ and we will identify τ and (a, b).

5 Parameter Estimation for Given Dynamic Specification

In most cases for given dynamic specification procedures for estimating the (real-valued) parameters $(\tau, \Sigma) = \theta$, the parameters are obtained by optimizing a criterion of (mis)fit of the system to the data. For the AR case, least-square-type procedures give numerically fast and statistically satisfactory estimators. Estimation in the ARMA case is more complicated, partly because the ε_{t-j} values are not directly observed.

The most important class of estimation procedures in this case are of the (Gaussian) maximum likelihood (ML) (or prediction error variance minimization) type. An example of this is the Whittle likelihood

$$
\begin{aligned}
L_T(\tau, \Sigma) \;=\; & \log\det\Sigma + (2\pi)^{-1}\mathrm{tr}\int_{-\pi}^{\pi}(k(e^{-i\lambda}) \\
& \times\Sigma\cdot k(e^{-i\lambda})^{*-1}I(\lambda)d\lambda
\end{aligned}
$$

where tr denotes "trace" and where

$$I(\lambda) = (2\pi)^{-1}\sum_{s=-T}^{T}\hat{\gamma}(s)e^{i\lambda s}$$

is the periodogram, with T being the sample size, and $\hat{\gamma}(s) = \frac{1}{T}\sum_{t=1}^{T-s} y_{t+s}y_t'$ for $s \geq 0$ and $\hat{\gamma}(s) = \hat{\gamma}(-s)'$ for $s < 0$ being the sample covariances. The Whittle estimators $\hat{\theta}_T = (\hat{\tau}_T, \hat{\Sigma}_T)$ are obtained from minimizing L_T over $U_\alpha \times \underline{\Sigma}$ where $\underline{\Sigma}$ is the set of all nonsingular $s \times s$ covariance matrices.

It is important to note that ML-type estimation contains a model reduction step, where the transfer function directly obtained from $I(\lambda)$ is approximated (within U_α) by $\hat{k}_T = \pi(\hat{\tau}_T)$.

Note that L_T depends on τ only via $k = \pi(\tau)$. For this reason, the Whittle likelihood can be considered as a function of k and Σ and estimation (theory) can partly be performed in a coordinate-free way. The data enter L_T only via the sample covariances.

One problem is that in general no explicit formula for the Whittle (or for ML) estimators $\hat{\tau}_T$ $\hat{\Sigma}_T$ exists. These estimators have to be obtained from an optimization algorithm. From a practical point of view, important questions in this context are the appropriate choice of initial estimators and problems arising from local minima.

A complete asymptotic theory (Dunsmuir and Hannan, 1976; Hannan and Deistler, 1988) for ML-type estimators is available now. The main results concern consistency for $\hat{k}_T = \pi(\hat{\tau}_T)$, $\hat{\Sigma}_T$, [and thus also for $\hat{\tau}_T = \varphi_\alpha(\hat{k}_T)$], generalized consistency for the case where the true system is not contained in the model class U_α and asymptotic normality.

6 Dynamic Specification

As has been stated already, in many applications the dynamic specification (i.e., α) is not known *a priori* and has to be determined from the data too. For the sake of simplicity, we consider estimation of p, where p stands for max (p, q) in the ARMA systems in equation (3). Then U_p is the subclass of U_A corresponding to (a, b) of degree less than or equal to p. An important class of estimators for p is obtained by minimizing information criteria of the form

$$A(p) = \log\det \hat{\Sigma}_T(p) + 2s^2 p \cdot \frac{c(T)}{T}$$

where $\hat{\Sigma}_T$ is (say) the Whittle estimator for Σ over $U_p \times \underline{\Sigma}$ and $c(T)$ is a prescribed (nonstochastic) function. Criteria of this type describe the tradeoff between the quality of fit of a system to the data [measured by $\log\det \hat{\Sigma}_T(p)$] and the complexity of the model (subclass) (described by the dimension of the parameter space, $2s^2p$). The particular tradeoff [described by $c(T)$] may be obtained, e.g., from maximizing the entropy or by Bayesian arguments (Akaike, 1977; Rissanen, 1983). The most important special cases are the AIC criterion defined by $c(T) = 2$ and the BIC criterion defined by $c(T) = \log T$. It can be shown (Hannan, 1980) that BIC gives consistent estimators for p, whereas AIC does not. However, AIC has other optimality properties.

7 Identification of Linear Systems – Alternative Approaches

For many applications linear systems with *time varying parameters* are important; then for instance the task may be to track the time path of slowly varying parameters or to detect structural breaks and to identify a system for each regime. Linear systems with time varying parameters are also often used as approximations for nonlinear systems. The parameter variation may be described by a stochastic model (e.g., as a random walk or by a Markov process with discrete states) or by deterministic functions (e.g., by polynomials or by harmonic functions). There is a substantial volume of literature now for such problems, particularly in connection with on-line estimation procedures. A novel approach to an asymptotic theory is given in Dahlhaus (1993).

Since *unstable linear systems* with stationary inputs in general have nonstationary outputs, the classical asymptotic theory with $\sqrt{T}$ consistency and Gaussian limiting distributions no longer applies in this case. Asymptotic results for unstable (single output) AR systems are given, e.g., in Lai and Wei (1983). A special form of instability, namely *unit roots* has attracted great attention in econometrics now, in particular in the context of cointegration (see, e.g., Engle and Granger, 1987).

Identification of systems under *feedback* poses a number of additional problems. Recently, the connection between *control and identification* has been studied in detail (Gevers, 1991).

Particularly in connection with control, alternative criteria for goodness of fit are of interest. In a certain sense ML-type criteria are prediction oriented, whereas in control, e.g., the uniform approximation of transfer functions (on the unit circle) is of interest. Recently, model reduction methods, such as truncation of balanced realizations and Hankel norm approximations have been considered for system identification (Aoki, 1987).

Errors-in-variables models, where inputs and outputs may be subject to noise have attained increased attention recently (Kalman, 1982; Deistler and Anderson, 1989). In this case the symmetric way of noise-modeling also allows for symmetric system models (Willems, 1986), where the classification of the variables into inputs and outputs and even the number of equations are not necessarily known *a priori*.

8 Nonlinear Systems

I like the statement that "nonlinear systems identification" is a word like "non-elephant zoology", because the class of nonlinear systems is so large when compared to the class of linear systems. Nonlinear systems identification is still a widely open area, despite the fact that there are a number of substantial results.

Consider, for instance, a parametric class of nonlinear systems of the form

$$y_t = f(y_{t-1}, \ldots, y_{t-p}, \varepsilon_t; \theta) \ , \quad \theta \in \Theta \subset \Re^p \ .$$

In such a case the same questions as in the linear case arise, namely:

- What can be said about identifiability and more generally about realization and parametrization?
- If estimation of the parameter θ, e.g., by nonlinear least squares, ML or the generalized method of moments is considered, what are the asymptotic properties of these procedures?
- Model selection problems such as estimation of p or discrimination between different model classes of course, in general, also occur, in most cases with a richer structure because of the richness of the class of nonlinear systems.

It is quite clear that for the first issue, a general theory can hardly be developed. In particular for the second issue, however, a general theory has been developed (see, e.g., Gallant, 1987; Pötscher and Prucha, 1991), using ideas analogous to the linear case. Note that estimation for linear systems is a nonlinear problem, thus in a sense there is no substantial difference to estimation for nonlinear systems. One main problem with this general nonlinear theory of estimation is that the assumptions may be quite difficult for verifying the specific cases at hand (e.g., identifiability conditions have to be checked then). For these reasons, special model classes have been investigated in more detail.

Before describing some examples of special model classes, let me just mention that nonparametric estimation procedures, such as nonparametric regression and estimation of higher order cumulant spectra, are of increasing importance.

A still rather general class of nonlinear systems admits a *Volterra series expansion of the form*

$$y_t = \mu + \sum_{j=0}^{\infty} K_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} K_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \cdots$$

where (ε_t) is Gaussian white noise (here for the sake of simplicity of notation no observed inputs are added) (see, e.g., Priestley, 1988). Another important class is *bilinear systems* (see e.g., Brockett, 1976; Granger and Andersen, 1978) which here are of the form

$$y_t + \sum_{j=1}^{p} A_j y_{t-j} = \sum_{j=0}^{q} B_j \varepsilon_{t-j} + \sum_{i=1}^{r}\sum_{j=1}^{s} C_{ij} y_{t-i}\varepsilon_{t-j} \ .$$

An example of a rather special but very useful class of nonlinear systems is *ARCH and GARCH models* (see, e.g., Engle and Bollerslev, 1986). The simplest ARCH model is of the form

$$y_t = a y_{t-1} + \varepsilon_t \qquad |a| < 1$$

where ε_t satisfies

$$E(\varepsilon_t / F_{t-1}) = 0 \ ; \quad E(\varepsilon_t^2 / F_{t-1}) = h_t = \omega + \beta \varepsilon_{t-1}^2 \ ;$$

$$\omega > 0 \ , \quad \beta \geq 0 \ .$$

Here F_t denotes the σ-algebra generated by ε_s, $s \leq t$. The disturbances ε_t are serially uncorrelated but not independent. For $\beta < 1$, we have

$$\sigma^2 = E\varepsilon_t^2 = \omega(1-\beta)^a$$

and

$$h_t - \sigma^2 = \beta(\varepsilon_{t-1}^2 - \sigma^2) \ .$$

The important feature is that (y_t) has time-dependent conditional means and variances, but is still stationary. In this way time series, for instance financial data, with time segments of high and low volatility can be modeled. For ARCH and GARCH models a rather complete theory of estimation has been developed.

Neural nets constitute an increasingly important class of nonlinear systems, also because of their computational advantages.

In addition to linear and nonlinear stochastic systems, special deterministic nonlinear systems, namely *chaotic systems*, have also recently attracted great attention for the modeling of time series (see, e.g., Liu *et al.*, 1992). One of the interesting aspects here is that trajectories from a chaotic system may have, in a certain respect, identical features to trajectories from a stochastic process; e.g., the deterministic system

$$y_{t+1} = 4y_t(1 - y_t) \quad , \quad y_0 \in (0,1)$$

generates trajectories, having "white noise properties" such as

$$\lim_{T\to\infty} \left[\sum_1^T (y_t - \overline{y})(y_{t-s} - \overline{y})\right] \left[\sum_1^T (y_t - \overline{y})^2\right]^{-1} = 0 \quad \text{for } s > 0 \ .$$

The main issues in this context are:

- Discrimination of "white chaos" from "truly stochastic" white noise or, more generally, of chaotic from stochastic systems, based on observations: discrimination between different classes of chaotic systems.
- Estimation of coefficients in chaotic systems.

9 Conclusion

System identification is a rich subject, both with regard to theories and methods. A rather complete theory for the linear case is available now. Identification of nonlinear systems is still a wide open area. Also in applications, linear models (still) dominate. In spite of the fact that there are already a great number of successful applications in many different areas, the real boom in applications is just about to start. In my opinion, important challenges for theories and methods in the future will come from applications. In many applications, the task of identification is not fully automatized in the sense of only applying a standard algorithm to data. However, the degree of automatization will increase in the future, also due to the use of knowledge-based methods.

References

Akaike, H., 1977, On Entropy Maximization Principle, in P.R. Krishnaiiah, ed., *Applications of Statistics* 27–41, North Holland, Amsterdam, The Netherlands.

Anderson, T.W., 1971, *The Statistical Analysis of Time Series*, Wiley, New York, USA.

Aoki, M., 1987, *State Space Modeling of Time Series*, Springer Verlag, Berlin, Germany.

Box, G.E.P., and Jenkins, G.M., 1970, *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, USA.

Brockett, R.W., 1976, Volterra Series and Geometric Control Theory, *Automatica* **12**:167–176.

Brockwell, P., and Davies, R., 1991, *Time Series: Theory and Methods*, Second Ed., Springer-Verlag, New York, USA.

Caines, E.P., 1988, *Linear Stochastic Systems*, Wiley, New York, USA.

Dahlhaus, R., 1993, Fitting Time Series Models to Nonstationary Processes, mimeo.

Deistler, M., 1989, Linear System Identification – A Survey, in J.C. Willems, ed., *From Data to Model*, Springer-Verlag, Berlin, Germany.

Deistler, M., and Anderson, B.D., 1989, Linear Dynamic Errors Invariables Models: Some Structure Theory, *J. Econometrics*, **41**:39–63.

Dunsmuir, W., and Hannan, E.J., 1976, Vector Linear Time Series Models, *Adv. Appl. Probab.* **8**:339–364.

Engle, R., and Bollerslev, T., 1986, Modelling the Persistence of Conditional Variances, *Econometric Reviews*, **5**:1–50.

Engle, R., and Granger, C.W.J., 1987, Co-integration and Error-correction: Representation, Estimation and Testing, *Econometrica* **55**:251–276

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

Gevers, M., 1991, Connecting Identification and Robust Control, A New Challenge, *Proc. 9th IFAC Symposium on Identification and Parameter Estimation*, **1**, 1, Budapest, Hungary.

Granger, C.W.J., and Andersen, A.P., 1978, *An Introduction to Bilinear Time Series Models*, Vandenhoeck und Ruprecht, Göttingen, The Netherlands.

Hannan, E.J., 1970, *Multiple Time Series*, Wiley, New York, USA.

Hannan, E.J., 1980, The Estimation of the Order of an ARMA Process, *Ann. Statist.* **8**:1071–1081.

Hannan, E.J., and Deistler M., 1988, *The Statistical Theory of Linear Systems*, Wiley, New York, USA.

Kalman, R.E., 1963, Mathematical Description of Linear Dynamical Systems, *SIAM J. Control* **1**:152–192.

Kalman, R.E., 1982, System Identification from Noisy Data, in A. Bednarak and L. Cesari, eds., *Dynamical Systems II, a University of Florida International Symposium*, Academic Press, New York, USA.

Lai, T.L., and Wei, C.Z., 1983, Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-squares Estimates of their Parameters, *J. Multivariate Anal.* **13**:1–23.

Liu, T., Granger, C.W.J., and Heller, W.P., 1992, Using the Correlation Exponent to Decide Whether an Economic Time Series is Chaotic, *J. of Applied Econometrics* **7**, Suppl., 525–540.

Ljung, L., 1987, *System Identification – Theory for the User*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA.

Pötscher, B.M., and Prucha, I.R., 1991, Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts, *Econometric Reviews* **10**:125–216.

Priestley, M.B., 1988, *Non-linear and Non-stationary Time Series Analysis*, Academic Press, London, UK.

Rissanen, J., 1983, Universal Prior for Parameters and Estimation by Minimum Description Length, *Ann. Statist.* **11**:416–431.

Söderström, T., and Stoica P., 1988, *System identification*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA.

Willems, J.C., 1986, From Time Series to Linear System, Part I: Finite Dimensional Linear Time Invariant Systems, *Automatica* **22**:561–580.

Rapporteur's Report

István Vályi

The invited speaker, Professor Rudolf Kalman was not able to come, and therefore one of the discussants, Professor Manfred Deistler – who is a coauthor with Professor Kalman in work on the subject – presented his paper under the title *A Survey on System Identification.*

In conclusion, with reference to Kalman's well-known militant views, he pointed out that various *assumptions* are inevitable for making analytical techniques work, but they *are by no means innocent*, and can play a decisive and occasionally arbitrary role in the outcomes. These include:

- assumptions on the noise structure, (like the problem of errors in variables, or in other terms, how the distance of the data from the model is measured);
- those implicit in the way and structure of sampling;
- and such aspects of model structure that remain open even after restrictions to finite-dimensional, linear models with stationary, ergodic, noise as the number of dependent variables (outputs).

The major points of Professor Deistler's presentation, complemented by those of the quite vivid discussion, are the following – as perceived by the rapporteur.

Identification techniques are used in a rather wide spectrum of applications, where different amounts of *a priori* knowledge are available to support the above. So, e.g., in signal processing there is usually no *a priori* information, while in natural sciences, in the case of testing theories, there may be a sufficient amount available.

Dynamics Systems Project, IIASA, Laxenburg, Austria

Under favorable conditions, these methods serve as a systematic way of incorporating intuition into scientific theories. Within this framework various statistical methods are at our disposal to identify causality.

Using discrete versus continuous-time models is closely related to the question of sampling. It appears that the former is quite widely used. The technical condition for this is that the high frequency spectrum could be neglected. In connection with the dilemma of using linear or nonlinear models, it was pointed out that many nonlinear applications are reported, but, as the latter is such a vast subject, from a theoretical point of view it can be nicknamed the field of "non-elephant zoology". One is faced with such choices in the case of semi-stability, or doubts of well-posedness. It must be stressed that nonlinearity and continuity are independent attributes. If both problems need to be considered, since each involve technical difficulties, it is advisable to tackle them separately.

A well-known example of nonlinearity-related problems arises in analyzing technological progress in economic theory. It is the phenomenon of limit cycles changing over time – to be modeled by nonlinear systems plus random errors. Technological innovation, however, can be represented in various other ways, like changes in parameter values, or in model structure. By introducing random shocks into the seemingly unfit linear models of economic cycles one can achieve keeping the system alive – while in the case where the real parts of the eigenvalues are less than one, the trajectories converge to a constant.

In the case of the necessity of nonlinear models, there are good compromises, like applying bilinear, threshold, or Volterra-models that are usually sufficiently rich to represent nonlinear features, and at the same time relatively easy to handle.

Linear models give often good descriptions locally in time and space, but become inadequate in the global sense. One theoretical tool to be used here is the "catastrophe" theory, which also offers solutions if nondifferentiability is an essential feature. Another method is the one based on the differential geometrical approach. In models involving stochastic differential equations, matringale theory allows us to relax conditions about the distributions of the random variables representing the errors.

Still, the major open problems in econometrics include: treating nonlinearities, and studying the statistic properties of nonlinear models, where errors may play many roles. Much attention has been recently given to microeconomics, where a large number of small units (like households) are analyzed, and attempts are made to arrive at qualitative conclusions, based on quantitative data. A remarkable approach here, with promising results, is the one using analogies to statistical mechanics, such as equivalents to the notion of entropy.

Current trends in identification show interest in conceptual issues and general paradigms decreasing, while more efforts are being devoted to special, detailed questions, and to improving the performance of concrete applications.

Related to growing attention to global issues, and IIASA's decision to make the subject its focus of interest, it was emphasized, that:

- *It would be a capital error to weaken the theoretical basis* and reduce mathematics using the complexity of the problem as an excuse. The challenge of the treatment of international problems should be met by maintaining high standards and rather stimulate new methodological achievements. ("Avoid using metalanguage only!")
- Another aspect worth considering is the question of interventionism. (Former President Bush: "I don't want social engineering.") The present unpopularity of the approach does not necessarily imply that IIASA should abandon it, perhaps it is better to pursue this great task. *The study of the combination of social phenomena and environment is an investment in the future.*
- Also, concerning IIASA's plans, *interdisciplinary work was pointed out as a tradition to be maintained.* Based on this an eventually important function of the institute could be to educate young researchers to work in an interdisciplinary setting.