Estimation of Factor Models by Realization-Based and Approximation Methods

Wolfgang Scherrer *
Institut für Ökonometrie, Operations Research und Systemtheorie
Technische Universität Wien
Argentinierstrasse 8/119, A-1040 Vienna, Austria
W.Scherrer@tuwien.ac.at

Christiaan Heij
Econometric Institute and Tinbergen Institute
Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
heij@few.eur.nl

Submitted to the Proceedings of the MTNS-98, Mathematical Theory of Networks and Systems, Padova, Italy, 1998

Econometric Institute Report 9831/A

Keywords
Linear system, system identification, model reduction, principal components, subspace method, behaviour.

Abstract
In this paper we discuss two methods for the estimation of linear dynamic factor models. The first method is behavioural in nature and consists of the least squares approximation of the observed data by means of a linear system. The second method is based on the statistical concept of principal components and uses subspace ideas from approximate realization theory. The two methods are compared by means of simulated data.

*Support by the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" Project P-11213-MAT is gratefully acknowledged.

1 Introduction

The essential ingredients of system identification consist of data, models, and methods to identify models from data. In this paper we restrict the attention to time series data and linear system models and we focus on identification methods within this setting. These methods can be classified in terms of the following specifications of the identification problem :

a. symmetric treatment of variables or not ?

b. equations for all the variables or not ?

c. explicit noise model or not ?

d. explicit identification criterion or not ?

e. which model parametrization ?

In Table 1 we summarize some possible specifications of some well-known identification methods, that is, maximum likelihood estimation of ARMA processes (see [4]), prediction error identification of input-output systems (see [7]), subspace identification methods (see [14]), along with the identification problem studied in this paper. That is, in this paper we make no *a priori* distinction between inputs and outputs and we do not require that all the variables are restricted by the system. We consider two identification methods within this setting. The first method uses an explicit identification criterion corresponding to (nonlinear) least squares approximation of the data (see [6, 10, 12]). The second method is based on the statistical concept of principal components (see [1]) and uses approximate realization algorithms in the spirit of subspace identification (see [8, 15]). Our analysis extends earlier results in [13].

	a	b	c	d	e
ARMA	Y	Y	Y	Y	polynomial
IO-PEI	N	N	Y	Y	polynomial
SSI	Y/N	Y/N	N	N	state space
FAI	Y	N	N	Y/N	behaviour

Table 1: Some popular specifications (a-e) for the identification methods ARMA (maximum likelihood estimation of stationary ARMA processes), IO-PEI (prediction error identification of input-output systems), SSI (subspace identification of stationary processes and input-output systems) and FAI (factor analysis identification as studied in this paper).

2 Factor Models

Assume that the data are generated by a linearly regular stationary process w with absolutely summable autocovariance function and with spectral density $S(e^{i\omega})$. We define a factor model of a q-dimensional process w as a decomposition $w = \hat{w} + \tilde{w}$ where the factor process $\hat{w}$ satisfies linear system restrictions and where the process $\tilde{w}$ is the error resulting from the approximation of w by the reduced process $\hat{w}$. Expressed in behavioural terms, $\hat{w} \in \mathcal{B}$ where $\mathcal{B} \subset (R^q)^Z$ is a linear, shift invariant set that is closed (in the topology of pointwise convergence). For background on behaviours we refer to [17, 18] and for contributions on factor models to [2, 3, 9, 16].

The complexity of a factor model is measured by the pair of integers (m, n), defined by the property that $\dim(\mathcal{B}_N) = mN + n$ where $\mathcal{B}_N$ denotes the restriction of the behaviour $\mathcal{B}$ to a time interval of (sufficiently large) length N. When expressed in terms of input-output systems, m is the number of inputs and n is the (minimal) number of states. We consider the following identification problem.

For given process w and complexity (m, n) , find a factor model $w = \hat{w} + \tilde{w}$ such that

1. *$\hat{w} \in \mathcal{B}$ where $\mathcal{B}$ has complexity (m, n), and*
2. *$\tilde{w}$ is small.*

In practice the process w is not known and the data w_N consists of a q-dimensional vector $w(t)$ observed on the time interval $t = 1, \cdots, N$. The problem then becomes to make a decomposition $w_N = \hat{w}_N + \tilde{w}_N$ of the data with $\hat{w}_N \in \mathcal{B}_N$ and with $\tilde{w}_N$ small.

3 Behavioural Least Squares

In behavioural least squares (BLS), the system $\mathcal{B}$ is determined by minimization of the mean squared error

$$e(w_N, \mathcal{B}) = \min\{\frac{1}{N}\sum_{t=1}^{N} \|w(t) - \hat{w}(t)\|^2; \hat{w}_N \in \mathcal{B}_N\} \tag{1}$$

This differs from conventional least squares methods in three respects. First, all the variables are approximated (whereas conventionally the outputs are regressed on the inputs). Second, the approximation $\hat{w}$ should satisfy all the system restrictions of $\mathcal{B}$ (whereas conventionally the restrictions are evaluated only locally in time). Third, the problem is formulated in behavioural terms and does not require any parametrization of linear systems. Of course, for computational reasons the systems must be parametrized in some way. The resulting minimization problem is highly non-linear. For topological aspects of this problem we refer to [5], and for a Gauss-Newton algorithm in terms of state space models to [11].

4 Principal Components

For fixed frequency ω let the eigenvalue decomposition of the spectral density S of the process w be given by $S(e^{i\omega}) = \sum_{j=1}^{q} \lambda_j(e^{i\omega})u_j(e^{i\omega})u_j^*(e^{-i\omega})$, with $\lambda_1 \geq \cdots \geq \lambda_q > 0$. Under certain regularity conditions the $q \times 1$ eigenfunctions u_j can be chosen to be analytic. Then the $q \times m$ transfer function $\hat{U}$ with columns $u_j, j = 1, \cdots, m$, and the $q \times (q-m)$ transfer function $\tilde{U}$ with columns $u_j, j = m+1, \cdots, q$, are well-defined, at least if $\lambda_m \neq \lambda_{m+1}$. The $q \times q$ matrix $U = [\hat{U}\ \tilde{U}]$ is unitary, that is, U^*U is the identity matrix. The principal component model of order m is defined by the decomposition $w = \hat{w}_{PC} + \tilde{w}_{PC}$, where

$$\hat{w}_{PC} = \hat{U}(\sigma)\hat{U}^*(\sigma^{-1})w,\ \tilde{w}_{PC} = \tilde{U}(\sigma)\tilde{U}^*(\sigma^{-1})w \tag{2}$$

It was proposed in [1] to define $\hat{w}_{PC}$ as the factor process. This process has rank m and satisfies the restriction

$$\hat{w}_{PC} \in \mathrm{im}(\hat{U}\hat{U}^*) = \mathrm{im}(\hat{U}) \tag{3}$$

This model is optimal in the sense that it minimizes the error

$$\mathrm{E}\|w(t) - \hat{w}(t)\|^2 = \mathrm{trace} \int_{-\pi}^{\pi} (I - \hat{U}\hat{U}^*)S(I - \hat{U}\hat{U}^*)d\omega \tag{4}$$

over all factor processes $\hat{w}$ that have a spectrum with rank at most m. However, in general $\hat{U}\hat{U}^*$ is not a rational function and this means that the system corresponding to this model is infinite dimensional. In order to obtain a model of complexity (m, n), in the next section we use subspace methods to reduce the dimension of the state space.

In practice, the spectrum S of the process is not known and has to be estimated from the observed data w_N. For our purposes, good results are obtained by estimating a sufficiently long (q-dimensional vector) autoregression for the data and defining the spectrum in terms of the estimated model.

5 Subspace Reduction

We now give an outline of two methods for reducing the state space that are improvements of the methods introduced in [13]. The methods are based on the fact that a behaviour of complexity (m, n) can be expressed as $\mathcal{B} = \mathrm{im}(\hat{G}_n(\sigma))$ where $\hat{G}_n$ is a $q \times m$ rational matrix function of McMillan degree n. Note that the matrix $\hat{U}$ in (3) is not rational in general. Therefore the main idea is to approximate $\hat{U}$ by a rational matrix function $\hat{G}$ of given McMillan degree n. The two methods, that we call the iterative relation method (IRM) and the projection factorization method (PFM), can be summarized as follows.

1. Estimate an AR(K) model $w(t) = \sum_{k=1}^{K} A_k w(t-k) + \varepsilon(t)$ (where $K < N$ is determined by the AIC criterion) and determine the corresponding spectrum $S(e^{i\omega}) = (\sum_{k=1}^{K} A_k e^{ik\omega})\Omega(\sum_{k=1}^{K} A_k^* e^{-ik\omega})$, where Ω is the estimated covariance matrix of ε.

2. Determine the $q \times m$ eigenvector matrix $\hat{U}$ of $S(e^{i\omega})$ at the frequencies $\omega_l = 2l\pi/L, l = 0, \cdots, L-1$ (with $L \leq N$ sufficiently large).

3. Determine a $(q \times m)$ rational matrix function $\hat{G}$ of McMillan degree n and define $\mathcal{B} = \mathrm{im}(\hat{G})$, where
IRM : $\hat{G}$ approximates $\hat{U}$,
PFM : $\hat{G}\hat{G}^*$ approximates $\hat{U}\hat{U}^*$.

Because of space limitations we will not describe all algorithmic details of IRM and PFM here. The ideas are as follows.

IRM is an iterative method. An initial model is estimated as follows. Let the rows of $\hat{U}$ be partitioned into two groups, giving an $m \times m$ invertible matrix $\hat{U}_2$ and a $(q-m) \times m$ matrix $\hat{U}_1$. The property $\hat{w}_{PC} \in \mathrm{im}(\hat{U})$ in (3) can then also be expressed in input-output form as $\hat{w}_1 = k(\sigma)\hat{w}_2$, where $k = \hat{U}_1\hat{U}_2^{-1}$ and where $\hat{w}_1$ (the $q-m$ outputs) and $\hat{w}_2$ (the m inputs) are obtained from $\hat{w}$ by a partitioning conforming to that of $\hat{U}$. For the given frequency data $k(e^{i\omega_l}) = \hat{U}_1(e^{i\omega_l})\hat{U}_2^{-1}(e^{i\omega_l}), \omega_l = 2l\pi/L$, $l = 0, \ldots, L-1$, a rational approximation $\hat{k}$ of McMillan degree n is computed by the subspace frequency domain identification method (SFDIM) proposed in [8]. Since $k(\sigma)$ is in general mixed causal/anti-causal, SFDIM has to be generalized along the lines of [15]. This procedure is applied for all possible (q over m) selections of inputs and outputs, and the initial model in IRM is the one that minimizes the following squared approximation error (this is the finite sample version of (4)).

$$\mathrm{trace}\ \frac{2\pi}{L}\sum_{l=0}^{L-1}(I - P(e^{i\omega_l}))S(e^{i\omega_l})(I - P(e^{i\omega_l})) \qquad (5)$$

Here $P = \hat{G}(\hat{G}^*\hat{G})^{-1}\hat{G}^*$ is the projection onto the behaviour $\mathcal{B} = \mathrm{im}(\hat{G})$, where $\hat{G}$ is partitioned conforming with $\hat{U}$ with $\hat{G}_1 = \hat{k}$, $\hat{G}_2 = I$. Now observe that an isometric matrix function with the property $\hat{w}_{PC} \in \mathrm{im}(\hat{U})$ in (3) is only defined up to right multiplication with an $m \times m$ unitary matrix function. For a given approximation $\hat{G}$, we determine unitary matrices O_l such that $\|\hat{U}(e^{i\omega_l})O_l - \hat{G}(e^{i\omega_l})\|$ is minimal and apply SFDIM to $\hat{U}(e^{i\omega_l})O_l$ to get an improved estimate $\hat{G}$. This is iterated until the error (5) does not improve significantly anymore.

PFM uses the projection operator $\hat{U}\hat{U}^*$ as its starting point. The discrete inverse FFT is applied to $\hat{U}(e^{i\omega_l})\hat{U}(e^{i\omega_l})^*$ to get the equivalent of an autocovariance function. Then a modified (time domain) subspace technique is used to determine a $q \times m$ rational transfer function $\hat{G}$ of McMillan degree n so that $\hat{G}\hat{G}^*$ is a good approximation of the projection $\hat{U}\hat{U}^*$. The main technical complication here is that $\hat{U}\hat{U}^*$ has rank $m < q$. Let the $q \times q$ transfer function G denote the estimate obtained by a variant of N4SID (see [14], using a different weighting scheme). Furthermore let $O \in R^{q\times m}$ be a unitary basis of the eigenspaces of the matrix $\int G^*Gd\omega$ corresponding to its m largest eigenvalues, then $\hat{G} = GO$.

6 Simulation

Some of the statistical properties of these algorithms were evaluated by a simulation study. As an example we present some results for the bivariate case. The spectrum of the two-dimensional process w is given by $S = \alpha(u_1u_1^*) + (1-\alpha)(u_2u_2^*)$, where u_1 and u_2 are rational stable 2×1 transfer functions of McMillan degrees 1 and 5 respectively. Both functions are scaled such that $\int u_i^*u_id\omega = 1$. For the simulations we consider three 'scenarios' for the data generating process, namely $\alpha = 0.9, \alpha = 0.5$ and $\alpha = 0.1$. For $\alpha = 0.9$ and $\alpha = 0.1$ the behaviour given by the principal components model (3) should be close to the behaviour $\text{im}(u_1)$ and $\text{im}(u_2)$ respectively. Therefore, in the first case a system with $n = 1$ should give a good description of the data, whereas in the second case it should be harder to find a good model with less than $n = 5$ states.

For each of the three scenarios we simulate 200 runs with samples of length $N = 512$. For every data sample we estimate factor models with complexities $m = 1$ and $n = 1, 3, 5$ by means of the methods IRM, PFM and BLS. For IRM and PFM we use $L = 32$ frequencies.

Table 2 gives some summary statistics. The methods BLS, IRM and PFM identify behaviours $\mathcal{B}$ with corresponding errors $\min\{\text{E}\|w(t) - \hat{w}(t)\|^2;\ \hat{w} \in \mathcal{B}\}$. This is equal to

$$\text{trace} \int_{-\pi}^{\pi} (I - P_{\mathcal{B}})S(I - P_{\mathcal{B}})d\omega \tag{6}$$

where S is the process spectrum and $P_{\mathcal{B}}$ is the orthogonal projection operator onto the behaviour $\mathcal{B}$ (see [12] for computational details). The means and standard deviations in Table 2 refer to this criterion. The results (and other simulation experiments) suggest that BLS gives a good estimate in general. However, for complex models BLS is quite time consuming and the faster algorithms IRM and PFM do a comparable job. On average, IRM gives better results than PFM.

α	n	Alg	Mean	St.D.	MNE	Best	Time
0.9							
	1	IRM	0.0752	0.0011	.0624	0.5	0.8
		PFM	0.0751	0.0010	.0608	1.0	0.5
		BLS	0.0736	0.0008	.0398	98.5	1.1
	3	IRM	0.0753	0.0014	.0637	9.0	0.9
		PFM	0.0761	0.0032	.0751	13.5	0.6
		BLS	0.0740	0.0016	.0464	77.5	3.4
	5	IRM	0.0756	0.0015	.0682	22.5	1.2
		PFM	0.0775	0.0041	.0948	14.5	0.7
		BLS	0.0749	0.0020	.0588	63.0	6.9
0.5							
	1	IRM	0.2972	0.0150	.2315	7.5	0.8
		PFM	0.3218	0.0449	.3337	2.5	0.5
		BLS	0.2878	0.0073	.1927	90.0	1.7
	3	IRM	0.2617	0.0080	.0845	61.0	1.1
		PFM	0.2665	0.0127	.1043	17.0	0.5
		BLS	0.2694	0.0129	.1164	22.0	7.3
	5	IRM	0.2544	0.0057	.0544	81.0	1.2
		PFM	0.2597	0.0099	.0761	12.0	0.6
		BLS	0.2671	0.0119	.1070	7.0	13.8
0.1							
	1	IRM	0.1234	0.0025	.6607	2.0	0.7
		PFM	0.1233	0.0022	.6585	1.5	0.5
		BLS	0.1197	0.0007	.6110	96.5	1.0
	3	IRM	0.0773	0.0010	.0403	44.5	0.9
		PFM	0.0775	0.0013	.0431	38.5	0.5
		BLS	0.0789	0.0053	.0616	17.0	4.6
	5	IRM	0.0764	0.0010	.0285	87.0	1.2
		PFM	0.0783	0.0034	.0535	7.0	0.6
		BLS	0.0785	0.0017	.0563	6.0	8.7

Table 2: Results for the simulations described in Section 6. The simulations consist of 200 runs with samples of length $N = 512$ of three data generating processes. The columns show the values of α (defining the data generating process), n (the state dimension of the estimated factor process), alg (the employed identification algorithm), mean (the mean of the approximation errors (6) over the 200 runs), st.d. (the standard deviation of these errors), mne (the mean of the normalized errors $(e-e_{PC})/e_{PC}$, where e is the error (6) and e_{PC} is the error (4) of the principal components model), best (the percentage of the 200 runs for which the considered algorithm has the smallest error of the three algorithms), and time (the mean computation time of the algorithms).

7 Conclusions

In this paper we considered two types of methods for the identification of dynamic factor models of given complexity (m, n). BLS amounts to a nonlinear least squares problem over the set of all behaviours with m inputs and n states. IRM and PFM approximate the principal component model (with m inputs) by a system with n states by means of subspace techniques. A simulation study suggests that BLS performs relatively well for models of low complexity, whereas for more complex models the subspace techniques might give better results.

References

[1] D.R. Brillinger. *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York, 1975.

[2] M. Deistler. Symmetric modeling in system identification. In H. Nijmeijer and J.M. Schumacher (eds.), *Three Decades of Mathematical System Theory*, 128-147, Springer, Berlin, 1989.

[3] J.F. Geweke. The dynamic factor analysis of economic time series models. In D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-economic Models*, 365-383, North Holland, 1977.

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

[5] C. Heij and W. Scherrer. Consistency of system identification by global total least squares. *Report* 9635, Econometric Institute, Erasmus University Rotterdam, 1996.

[6] C. Heij, W. Scherrer and M. Deistler. System identification by dynamic factor models. *SIAM Journal on Control and Optimization* 35:1924-1951, 1997.

[7] L. Ljung. *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[8] T. McKelvey. *Identification of State Space Models from Time and Frequency Data.* Dissertation no. 380, Linköping Studies in Science and Technology, Linköping, 1995.

[9] G. Picci and S. Pinzoni. Dynamic factor analysis models for stationary processes. *IMA Journal of Mathematical Control and Information* 3:185-210, 1986.

[10] B. Roorda. *Global Total Least Squares.* Tinbergen Institute Research Series no. 88, Thesis Publishers, Amsterdam, 1995.

[11] B. Roorda. Algorithms for global total least squares modelling of finite multivariable time series. *Automatica* 31:391-404, 1995.

[12] B. Roorda and C. Heij. Global total least squares modelling of multivariable time series. *IEEE Transactions on Automatic Control* 40:50-63, 1995.

[13] W. Scherrer and C. Heij. Identification of factor models by behavioural and subspace methods. *Systems and Control Letters* 32:335-344, 1997.

[14] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems*. Kluwer, Boston, 1996.

[15] P. Van Overschee, B. De Moor, W. Dehandschutter and J. Swevers. A subspace algorithm for the identification of discrete time frequency domain power spectra. *Automatica* 33:2147-2157, 1997.

[16] J.H. Van Schuppen. Stochastic realization problems. In H. Nijmeijer and J.M. Schumacher (eds.), *Three Decades of Mathematical System Theory*, 480-523, Springer, Berlin, 1989.

[17] J.C. Willems. From time series to linear system, part I: Finite dimensional linear time invariant systems. *Automatica* 22:561-580, 1986.

[18] J.C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control* 36:259-294, 1991.