

# Behavioural Approximation of Stochastic Processes by Rank Reduced Spectra

C. Heij \*      W. Scherrer †

## Abstract

Behaviours provide an elegant, parameter free characterization of deterministic systems. We discuss a possible application of behaviours in the approximation of stochastic systems. This can be seen as an extension to the dynamic case of the well-known static factor analysis model. An essential difference is that we see modelling primarily as a matter of process approximation, not as a method to recover the true data generating process. In particular we see "noise properties" as a kind of prior model assumption that can be compared with the resulting quality of the process approximation.

## Keywords:

Linear systems, stationary processes, behaviours, factor analysis, least squares.

This paper has been submitted to 35th IEEE Conference on Decision and Control, 1996.

---

\*Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, phone +31-10-4081277, fax +31-10-4527746, e-mail [heij@ect.few.eur.nl](mailto:heij@ect.few.eur.nl).

†Institut für Ökonometrie, Operations Research und Systemtheorie, Technische Universität Wien, Argentinierstrasse 8, A-1040 Wien, Austria, phone +43-1-58801-4405, fax +43-1-5054524, e-mail [wo\\_sch@e119ws1.tuwien.ac.at](mailto:wo_sch@e119ws1.tuwien.ac.at).

# 1 Open Symmetric Models

In this brief, expository paper we are concerned with a new modelling strategy. In order to give a clear demarcation of the proposed approach from existing methodologies we classify them according to the following two criteria.

- closed - open
- symmetric - asymmetric.

We call a model closed if it gives a complete description of the relevant properties of all the model variables. It is open if these properties are not fully specified for all the model variables. A model is called symmetric if all the model variables are restricted in a similar way, asymmetric if some of the variables are treated different from others.

These are of course somewhat imprecise notions, but they are helpful in describing the main differences between various modelling approaches. In our case we are interested in linear dynamical systems that evolve in discrete time, and we will distinguish deterministic from stochastic models. As properties of interest we consider the dynamical relations between the system variables, that is, (linear) dynamical equations in the deterministic case and dynamical correlations in the stochastic case. Together with the above two criteria we can then distinguish eight types of models.

- **Deterministic models**

- closed symmetric : autonomous difference equations
- closed asymmetric : unusual
- open symmetric : behaviours
- open asymmetric : input - output systems.

- **Stochastic models**

- closed symmetric : stationary processes
- closed asymmetric : unusual
- open symmetric : factor models
- open asymmetric : input - output systems.

The mainstream approaches choose either for closed and symmetric models or for open and asymmetric ones. This can be explained as follows. Closed symmetric models are nice mathematical objects that serve the study of well-understood dynamical processes in, for example, physics. In other fields, the aim is often not so much to give a closed form description of the future evolution of the process. For example, in engineering one may wish to keep unexplained some parts of the system environment that influence the dynamics. In other fields, for example in econometrics, one only has partial information so that a closed system description is beyond our abilities. The usual step is then to

build conditional models that explain the behaviour of some of the system variables (the outputs) in terms of other variables (the inputs) that are left unexplained by the model.

What about the other classes of models ? As far as linear systems are concerned, we believe that closed asymmetric models will not be used so much. This is of course different for non-linear systems, for example hybrid systems where some of the variables are discrete valued and others not. In this paper, however, we restrict our attention to linear systems and consider open symmetric models. In our opinion this class of models is of considerable interest, both from a conceptual point of view and with regard to possible applications. It has relevance in all those situations where

- the system environment falls outside the scope of the model or one has incomplete knowledge, so that a closed model is not appropriate;
- there are no natural conditioning variables, so that an asymmetric model is not appropriate.

Several methods for the open symmetric modelling of systems have been developed. In Section 2 we integrate two of these approaches, namely the behavioural modelling of deterministic dynamical systems and factor models for stochastic static systems. We discuss some modelling questions related with this new class of dynamic factor models in Section 3. For a more extensive treatment of these issues we refer to [7], and for further background on system behaviours to [12], [13].

## 2 Dynamic Factor Models

We denote the observed variables by the  $q$ -dimensional real-valued vector  $w$ . In its most general form, the static factor analysis model is of the form

$$w = f + e \tag{1}$$

where  $f$  is the factor part and  $e$  the error part. The factor part satisfies linear restrictions, say  $Af = 0$  where  $A$  is a  $p \times q$  matrix with full row rank  $p \leq q$ . Let  $\mathcal{A} := \ker(A)$  and let  $F$  be a  $q \times (q - p)$  matrix with  $\text{im}(F) = \mathcal{A}$ , then we can describe this model also as

$$w = Fv + e \tag{2}$$

This model states that the observations can be approximated by a structural part within the subspace  $\mathcal{A}$  up to an additive error  $e$ . This is clearly an open symmetric model. In the literature several versions of this model have been investigated. A common assumption is that all involved variables are random and that the error term satisfies additional properties. In particular, it is usually assumed that  $e$  and  $v$  are uncorrelated and that also all the components of  $e$  are mutually uncorrelated. In terms of the covariance matrices  $\Sigma$  of  $w$ ,  $\Sigma_f$  of  $f$  and the diagonal covariance matrix  $\Sigma_e$  of  $e$  the factor model can be expressed as

$$\Sigma = \Sigma_f + \Sigma_e \tag{3}$$

Models of this type have been used in statistics, psychometrics and econometrics for a long time, and the associated structure theory, the so-called Frisch problem, has received attention in the systems community, see [1], [2], [10]. Models of this kind are relevant when the observed outcomes can be seen as the result of a limited number of underlying factors that are contaminated by independent errors. However, the independence assumption imposes very strong restrictions on the method of observation. In our approach we will refrain from such restrictions on the error part and instead we will focus on the approximation interpretation of factor models.

The definition of dynamic factor models that we will use in this paper is inspired by the decomposition (1). In the literature several models of this kind have been proposed that differ mainly in the assumptions on the noise process, see for example [4], [5], [9], [3] and [11]. In this paper we impose less restrictions on the noise process than is customary, and our aim is to find a good compromise between the complexity and the goodness of fit of factor models. Our definition of factor models is based on the behavioural approach to system modelling, as we will make more explicit now.

Let  $w$  be a  $q$ -dimensional stochastic process that evolves in discrete time and let  $\Sigma$  be its spectrum. A dynamic factor model corresponds to a decomposition  $w = f + e$  where the factor part  $f$  satisfies linear deterministic restrictions. More in particular, we assume that there exist  $1 \times q$  polynomial matrices  $a(z)$  in the time shift operator  $z$  such that  $a(z)f = 0$  holds true (almost surely). If the process  $f$  is assumed to be stationary then the set of all annihilating polynomial matrices forms a module. In this case let  $A(z)$  be a  $p \times q$  polynomial matrix with independent rows that span this module. If the process  $f$  has spectrum  $\Sigma_f$ , then this module coincides with the left polynomial kernel of  $\Sigma_f$

and  $\mathcal{B} := \ker(A(z))$  coincides with the image of  $\Sigma_f$  as subspace of all the  $q$ -dimensional discrete time sequences. We call  $\mathcal{B}$  the behaviour of the model, that is, the linear space that contains (almost all) realizations of the factor process  $f$ . This space coincides with a linear, finite dimensional system with  $p$  outputs and  $q - p$  inputs, see [12]. If this system is controllable then it can be represented in image form, that is,  $\mathcal{B} = \text{im}(F(z))$  for some  $q \times (q - p)$  polynomial matrix  $F(z)$ . Then we can write the process decomposition (1) in the form

$$w = F(z)v + e \quad (4)$$

In spectral terms this says that the observed spectrum  $\Sigma$  can be approximated by the rank reduced spectrum  $\Sigma_f = F \Sigma_v F^*$  up to an error process  $e$ . If we impose the condition that the processes  $f$  and  $e$  are uncorrelated, then  $\Sigma = \Sigma_f + \Sigma_e$  so that the error spectrum  $\Sigma_e$  measures the error in reducing the rank of the observed spectrum.

We now give a more formal definition of the above ideas. The behaviour of a stochastic process  $f$  is defined as the smallest linear, finite dimensional system that contains almost all process realizations, that is, it is the behaviour  $\mathcal{B}(f)$  with the following properties ( $\mathcal{B}$  denotes an arbitrary linear, finite dimensional system).

- $P\{f \in \mathcal{B}(f)\} = 1$
- $P\{f \in \mathcal{B}\} = 1 \implies \mathcal{B}(f) \subset \mathcal{B}$

**Proposition 1** *For every stochastic process the behaviour is well-defined.*

A behaviour is called non-trivial if it imposes restrictions, that is, if there are discrete time sequences that do not belong to the behaviour. A dynamic factor model is defined as follows.

**Definition 1** *A dynamic factor model of a process  $w$  is a decomposition of the form  $w = f + e$  where the factor process  $f$  has non-trivial behaviour.*

We will call the behaviour of  $f$  also the behaviour of the factor model. In order to simplify the analysis we make the following assumptions.

### Assumptions

- **A1** The processes  $w$ ,  $f$  and  $e$  are jointly weakly stationary, with zero mean and finite second order moments.
- **A2** The observed process  $w$  is purely nondeterministic and has full rank.
- **A3** The processes  $w$ ,  $f$  and  $e$  have absolutely summable Wold representations.

Assumption A1 is imposed for convenience, so that the usual tools of time series analysis and linear systems theory can be applied. Assumption A2 is a regularity condition. The full rank assumption implies that the observed process can not be modelled by a factor model with error zero. The process becomes purely nondeterministic if possible

deterministic components have been removed in a preliminary analysis, which is a usual assumption in time series analysis. It implies the existence of a moving average Wold representation. Finally, assumption A3 is imposed only for technical reasons. It implies that the spectral densities of the processes exist as bounded functions on the unit circle.

**Proposition 2** *Under assumption A3, the behaviour of factor models is controllable.*

This means that the model can be expressed as in (4). Alternatively, it can be written in state space form as

$$w_t = Cx_t + Dv_t + e_t, \quad x_{t+1} = Ax_t + Bv_t \quad (5)$$

This is a direct generalization of the static model (2). The state variables  $x$  make the dynamical effect of the driving factors  $v$  on the observed process  $w$  explicit.

### 3 Modelling Questions

For simplicity we will assume that complete information on the observed process  $w$  is available, in the sense that we know the process spectral density  $\Sigma$ . The question then becomes how to choose factor models for this process. In practice the only available information will often consist of an observed time series, and for the related identification questions we refer to [7], [6].

Borrowing terminology from Kalman [8], an interesting feature of dynamic factor models is that we can get no models without prejudice.

**Proposition 3** *For every behaviour  $\mathcal{B}$  there exists a factor model of the observed process  $w$  that has behaviour  $\mathcal{B}$ .*

So we have to impose prior restrictions in order to obtain models. Stated otherwise, our model formulation makes the choices explicit that are of a more hidden nature in modelling strategies that give unique models. The prior restrictions that we will consider here are related to structural requirements and to the complexity and goodness of fit of the models. We define the free factor scheme as the one that imposes no structural restrictions in addition to the one formulated in Definition 1. Further the following specifications are of interest.

**Definition 2** *A factor model (1) is called observable if  $f$  and  $e$  are linear functions of the process  $w$ . It is called orthogonal if the processes  $f$  and  $e$  are mutually uncorrelated.*

The result in Proposition 3 continues to hold for observable and orthogonal models. That is, for every behaviour there exist observable and orthogonal factor models. We base our choice of models on their explanatory power, in terms of their complexity and goodness of fit. We define the complexity of a factor model as the pair  $(m, n)$ , where  $m$  is the (minimal) number of driving forces  $v$  and  $n$  the (minimal) number of state variables in the state space representation (5). A model is less complex the smaller  $m$  and  $n$  are, that is, the complexity is inversely proportional to the degree of restrictiveness of the factor behaviour. The error of a factor model is measured in terms of the magnitude of the error process  $e$ . As norm we take the total sum of squares, that is,  $[E\|e_t\|^2]^{1/2} = \|\Sigma_e^{1/2}\|_2$ , where  $E$  denotes expectation,  $\|\cdot\|$  the Euclidean norm,  $\|\cdot\|_2$  the  $L_2$ -norm, and where  $\Sigma_e^{1/2}$  is a spectral factor of the error spectrum. Of course this choice of norm is somewhat arbitrary, but it has the virtue that it is symmetric in the model variables. This is in line with the symmetry of factor models. In practice one might consider prefiltering of the process to obtain a weighted least squares criterium that gives more importance to the errors in certain variables or in certain frequency bands.

Within this setting we prefer models that are less complex and that have better fit. For a given process spectrum  $\Sigma$ , the following modelling questions are of interest.

#### Questions

- **Q1** Given the behaviour  $\mathcal{B}$ , minimize the error  $\|\Sigma_e^{1/2}\|_2$  over all factor models with behaviour  $\mathcal{B}$ .

- **Q2** Given the complexity  $(m, n)$ , minimize the error  $\|\Sigma_e^{1/2}\|_2$  over all factor models with complexity  $(m, n)$ .

The solution of Question 1 is straightforward, and we refer to [7]. We denote the minimal error by  $\varepsilon_\Sigma(\mathcal{B})$ . It turns out that observability can be imposed without cost, but that the condition of orthogonality will in general lead to larger errors. Question 2 is much harder and has only been partly solved. We limit ourselves to some general aspects, for further details we refer again to [7].

Question 2 involves the minimization of the error  $\varepsilon_\Sigma(\mathcal{B})$  over the set  $\mathbf{B}(m, n)$  of all controllable linear systems with  $m$  inputs and  $n$  states. We denote by  $\overline{\mathbf{B}}(m, n) := \bigcup_{k=1}^n \mathbf{B}(m, k)$  the set of all controllable linear systems with  $m$  inputs and at most  $n$  states, and by  $\mathbf{B} := \bigcup_{m=0}^q \bigcup_{n=0}^\infty \mathbf{B}(m, n)$  the set of all controllable linear systems. On this model class we define the gap by  $d(\mathcal{B}_1, \mathcal{B}_2) := \|P_1 - P_2\|_\infty$ , where  $P_i$  is the orthogonal projection operator onto the set of square summable time series in the behaviour  $\mathcal{B}_i$ ,  $i = 1, 2$ , and where  $\|\cdot\|_\infty$  denotes the supremum norm. The gap defines a metric on  $\mathbf{B}$ . We summarize some properties that are of relevance in solving Question 2.

#### Proposition 4

- (i) Under assumption A2, the error  $\varepsilon_\Sigma(\mathcal{B})$  is continuous in  $\mathcal{B}$ .
- (ii) For  $n > 0$  and  $m < q$  the set  $\mathbf{B}(m, n)$  is neither open nor closed in  $\mathbf{B}$ .
- (iii) The set  $\overline{\mathbf{B}}(m, n)$  is the closure of  $\mathbf{B}(m, n)$  in  $\mathbf{B}$ .
- (iv) For  $n > 0$  and  $m < q$  the sets  $\mathbf{B}$  and  $\overline{\mathbf{B}}(m, n)$  are not compact.

Note that the case  $m = q$  is not of interest, as this corresponds to the trivial behaviour, and that the case  $n = 0$  is also of little interest as this concerns only static models. So, the analysis of Question 2 may seem somewhat complicated because of the non-compactness of the model class. However, this complication vanishes under a weak regularity condition, as follows. For given process spectrum  $\Sigma$  and given number of driving factors  $m < q$  let  $\varepsilon_n := \inf\{\varepsilon_\Sigma(\mathcal{B}); \mathcal{B} \in \mathbf{B}(m, n)\}$  be the infimum of attainable errors for models of complexity  $(m, n)$ . Then there holds that  $\varepsilon_{n-1} \geq \varepsilon_n$ , but strict inequality need not hold true. The regularity condition requires that the process spectrum is such that this inequality holds strictly. That is, it is assumed that the specified complexity  $(m, n)$  is such that the optimal fit can not be approximated arbitrarily well by models of lower complexity  $(m, n')$  with  $n' < n$ . Indeed, if this were the case then the complexity was not chosen efficiently, as the complexity could be reduced without any substantial increase of the error. For given  $\varepsilon > 0$  we define the set  $\mathbf{B}_\varepsilon(m, n)$  as the set of models in  $\mathbf{B}(m, n)$  that have error at most  $\varepsilon$ , that is, with  $\varepsilon_\Sigma(\mathcal{B}) \leq \varepsilon$ .

**Proposition 5** *The set  $\mathbf{B}_\varepsilon(m, n)$  is compact for all  $\varepsilon_n \leq \varepsilon < \varepsilon_{n-1}$ .*

In the regular case the set  $\mathbf{B}_\varepsilon(m, n)$  is non-empty, and the above result helps to solve the minimization connected with Question 2, see [7].



## 4 Conclusion

In this paper we introduced a new class of dynamic factor models for the approximation of stationary processes. The approximation, that is, the factor part satisfies deterministic dynamical relations. The realizations of the factor process belong to a behaviour, that is, to a linear system with less inputs than the number of observed variables. The approximation can also be seen as a rank reduction of the spectral density function of the observed process. This modelling approach can be characterized as open, in the sense that the dynamics is only partially specified, and symmetric, in the sense that the variables are treated all alike. We discussed some modelling questions related with the problem of minimizing the error under restrictions on the allowed model complexity.

## References

- [1] M. Deistler, Symmetric modelling in system identification. In H. Nijmeijer and J.M. Schumacher, eds., *Three Decades of Mathematical System Theory*, Lecture notes in control and information sciences vol. 135, Springer, 1989, pp. 128-147.
- [2] M. Deistler and W. Scherrer, Identification of linear systems from noisy data. In D. Brillinger et al., eds., *New Directions in Time Series Analysis*, part II, IMA vol. 46, Springer, 1992, pp. 21-42.
- [3] W.A. Fuller, *Measurement Error Models*. Wiley, 1987.
- [4] 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*, North Holland, 1977, pp. 365-383.
- [5] S. Haykin, ed., *Array Signal Processing*, Englewood Cliffs, New York, 1985.
- [6] C. Heij and W. Scherrer, Consistency of global total least squares modelling. *Manuscript*, 1995.
- [7] C. Heij, W. Scherrer and M. Deistler, System identification by dynamic factor models. *Report 9501/A*, Econometric Institute, Erasmus University Rotterdam, 1995. Submitted.
- [8] R.E. Kalman, A theory for the identification of linear relations. In H. Brezis et al., eds., *Colloques Lions*, 1989.
- [9] E. Nowak, Global identification of dynamic shock-error models. *Journal of Econometrics* 27, 1985, pp. 211-219.
- [10] J.H. van Schuppen, Stochastic realization problems. In H. Nijmeijer and J.M. Schumacher, eds., *Three Decades of Mathematical System Theory*, Lecture notes in control and information sciences vol. 135, Springer, 1989, pp. 480-523.
- [11] J.K. Tugnait, Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control* 37, 1992, pp. 476-485.
- [12] J.C. Willems, From time series to linear system, part I: Finite dimensional linear time invariant systems. *Automatica* 22, 1986, pp. 561-580.
- [13] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems. *IEEE Transactions on Automatic Control* 36, 1991, pp. 259-294.