COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 7, No. 2, pp. 153-166, 2007

003

MODELLING HIGH-DIMENSIONAL TIME SERIES BY GENERALIZED LINEAR DYNAMIC FACTOR MODELS: AN INTRODUCTORY SURVEY

MANFRED DEISTLER* AND CHRISTIANE ZINNER*

Abstract. Factor models are used to condense high dimensional data consisting of many variables into a much smaller number of factors. Here we present an introductory survey to factor models for time series, where the factors represent the comovement between the single time series. Principal component analysis, linear dynamic factor models with idiosyncratic noise and generalized linear dynamic factor models are introduced and structural properties, such as identifiability, as well as estimation are discussed.

1. Introduction. Factor analysis has been developed by psychologists for measurement of intelligence in the beginning of the twentieth century. In particular Burt and Spearman, observing that in tests of mental ability of a person, the scores on different items tended to be correlated, developed the hypothesis of a common latent factor, called general intelligence, [6], [19]. In the 1930s, Thurstone and others proposed a more general model allowing for more than one common factor, representing different mental abilities, [22]. In general, the motivation for the use of factor models is compression of the information contained in a high dimensional data vector into a small number of factors and the idea of underlying latent nonobserved variables influencing the observations.

Whereas the initial approach to factor analysis was oriented to data originating from independent, identically distributed random variables and consisted in dimension reduction in the cross-sectional dimension (i.e. the number of variables), the idea has been further generalized to modelling of multivariate time series, thus compressing information in the cross-sectional and the time dimension. This idea has been pursued rather independently in a number of areas, such as signal processing, [5], or econometrics, [15], [17], [10].

The idea of factor analysis is of particular importance, if the cross-sectional dimension, n say, is large (in relation to sample size T), where the so called "curse of dimensionality" occurs. "Conventional" time series modelling, e.g. by autoregressive models, leads to a parameter space of dimension proportional to n^2, thus the com-

*Institute for Mathematical Methods in Economics, Research Unit: Econometrics and System Theory, University of Technology Vienna, Argentinierstr. 8 / 105-2, A-1040 Vienna, Austria. E-mail: Manfred.Deistler@tuwien.ac.at.

plexity of the model class shows quadratic dependence on n, whereas the number of data points, for fixed T is linear in n. Factor models are used to mitigate this curse of dimensionality.

The basic, common equation for all different kinds of factor models considered here is of the form

$$x_t = \Lambda(z)\xi_t + u_t = \chi_t + u_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where x_t is the n-dimensional vector of observations, ξ_t is the r-dimensional factor, u_t is the n-dimensional noise and the transfer function $\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j,\ \Lambda_j \in \mathbb{R}^{n\times r}$ is called the factor loading matrix; throughout we assume $\sum_{j=-\infty}^{\infty} |j|\|\Lambda_j\| < \infty$, where $\|\ \|$ denotes a norm. $\chi_t = \Lambda(z)\xi_t$ is called the common component or the latent variable. Here z is used both for a complex variable and for the backward shift on $\mathbb{Z}$.

Throughout we assume the following:

- $\mathrm{E}\xi_t = 0,\ \mathrm{E}u_t = 0$ for all $t \in \mathbb{Z}$.
- $\mathrm{E}\xi_t u_s' = 0$ for all $s,\ t \in \mathbb{Z}$, where $'$ denotes the transpose.
- (ξ_t) and (u_t) are wide sense stationary and (linearly) regular with covariances $\gamma_\xi(s) = \mathrm{E}\xi_t\xi_{t+s}'$ and $\gamma_u(s) = \mathrm{E}u_t u_{t+s}'$ satisfying

$$\sum_{s=-\infty}^{\infty} |s|\|\gamma_\xi(s)\| < \infty,\ \sum_{s=-\infty}^{\infty} |s|\|\gamma_u(s)\| < \infty. \tag{2}$$

- The spectral density f_χ of χ_t has rank r for all $\lambda \in [-\pi, \pi]$.

Then the spectral densities f_χ of (χ_t) and f_u of (u_t) exist as uniform limits of trigonometric polynomials and because (ξ_t) and (u_t) are uncorrelated we have using an obvious notation

$$f_x(\lambda) = \Lambda(e^{-i\lambda}) f_\xi(\lambda) \Lambda(e^{-i\lambda})^* + f_u(\lambda), \tag{3}$$

where $*$ denotes the conjugate transpose.

The third assumption above implies (see [5]) that, for the case when all eigenvalues are distinct, the eigenvalues of the spectral densities and suitably normalized eigenvectors have an analogous summability property as in (2).

A special case often considered occurs when $\Lambda(z) = \Lambda$ is constant and (ξ_t) and (u_t) and thus (x_t) are white noise. In this case (1) is called *static* and the variance matrix of x_t, Σ_x, is, using an obvious notation, of the form:

$$\Sigma_x = \Lambda\Sigma_\xi\Lambda' + \Sigma_u. \tag{4}$$

If $\Lambda(z) = \Lambda$ is constant, but (ξ_t) and (u_t) are not necessarily white noise, the model is called *quasi-static*.

The assumptions imposed so far do not determine a reasonable model class, in the sense that for given f_x, or Σ_x respectively, too many models would be possible, see [18]. Thus, in order to obtain reasonable model classes, further assumptions have to be imposed. This leads to principal component models, linear factor models with idiosyncratic noise and generalized linear factor models considered in this paper. For these model classes we are interested in

1. Estimation (of parametrized versions) of $\Lambda(z)$, f_ξ and f_u,
2. Estimation of the factors ξ_t and of the latent variables $\chi_t = \Lambda(z)\xi_t$,
3. Forecasting.

Proceeding to estimation in the narrow sense, problems of the structure of such models, in particular of identifiability have to be discussed.

2. Principal Component Analysis. The aim of principal component analysis (PCA) is to approximate the n-dimensional observed process (x_t) by a filtered version of itself, whose spectral density is of reduced rank r, such that the variance of the residuals is minimized, (see [5]). Hence, the additional assumption in the PCA-model is, that $\xi_t = C(z)x_t$ holds and that the $r \times n$ filter $C(z)$ and the $n \times r$ filter $\Lambda(z)$ are defined such that

$$\operatorname{tr}(\mathrm{E}u_t u_t') = \operatorname{tr}\Big(\mathrm{E}\big(x_t - \Lambda(z)C(z)x_t\big)\big(x_t - \Lambda(z)C(z)x_t\big)'\Big) \tag{5}$$

is minimized for fixed r, where tr denotes the trace.

Let us first consider the static case, where $\Lambda(z) = \Lambda$ and $C(z) = C$ are constant matrices. The solution of the minimization problem (5) is then obtained from the canonical representation of Σ_x, decomposed as

$$\Sigma_x = O_1\Omega_1 O_1' + O_2\Omega_2 O_2' = \Sigma_\chi + \Sigma_u, \tag{6}$$

where Ω_1 and Ω_2 denote the diagonal matrices consisting of the r largest and $(n-r)$ smallest eigenvalues of Σ_x, respectively, arranged in decreasing order and $O_1 = (o_1 \dots o_r)$ and $O_2 = (o_{r+1} \dots o_n)$ are the $(n \times r)$- and $n \times (n-r)$-dimensional matrices, respectively, of corresponding eigenvectors. Throughout we assume that all eigenvalues are distinct.

Then decomposition (6) is unique in the sense that Σ_χ and Σ_u are unique for given Σ_x and r. However this is not true for Λ and ξ_t, since post-multiplying Λ with a nonsingular matrix P and pre-multiplying ξ_t with its inverse P^{-1} yields the same χ_t. By making the special choice

$$\xi_t = O_1'x_t, \quad \Lambda = O_1, \quad \Sigma_\xi = \Omega_1, \quad u_t = O_2O_2'x_t. \tag{7}$$

we obtain the PCA-model. Here the j-th factor, $(o_j' x_t), j \leq r$, is called the j-th principal component of (x_t).

If Λ is still constant, but (ξ_t) and (u_t) are allowed to be serially correlated, then we call the PCA-model (7) quasi static.

In the dynamic case, where $\Lambda(z)$ and $C(z)$ are in general two-sided filters, the minimization of (5) is solved by the eigenvalue decomposition of the spectral density f_x,

$$f_x(\lambda) = O_1(e^{-i\lambda})\Omega_1(\lambda)O_1(e^{-i\lambda})^* + O_2(e^{-i\lambda})\Omega_2(\lambda)O_2(e^{-i\lambda})^* = f_\chi(\lambda) + f_u(\lambda), \tag{8}$$

where Ω_1, Ω_2, O_1 and O_2 are defined analogously to the static case; however, they are now functions of the frequency λ. Under our assumptions we obtain the dynamic PCA-model by

$$\xi_t = O_1(z)^* x_t, \quad \Lambda = O_1(z), \quad f_\xi = \Omega_1(\lambda), \quad u_t = O_2(z)O_2(z)^* x_t. \tag{9}$$

Now $(o_j(z)^* x_t)\,, j \leq r$, is called the j-th dynamic principal component or principal component series of (x_t).

For estimation of the PCA-model analog estimators are used: The population second moments, Σ_x and f_x, are replaced by their estimators - typically the sample covariance and a non-parametric estimator of the spectral density. As can be shown under general assumptions, consistency of these estimators together with the assumption that all eigenvalues are distinct, which implies that the eigenvalues and suitably normalized eigenvectors are continuous functions of the original matrices, yield consistent estimators of Λ, Σ_ξ and Σ_u and $\Lambda(z)$, f_ξ and f_u respectively.

For the PCA-model the number of factors r is not intrinsic in the sense, that it is not a property of f_x. By the choice of r, the degree of dimension reduction and, as a trade-off, the quality of approximation are determined. In dynamic PCA dimension reduction in the time dimension is performed by introducing a finite dimensional parametrization. Note, however, that even for rational f_x, the matrices on the right hand side of (8) are not necessarily rational.

3. Factor models with idiosyncratic noise. Here, in addition to the general assumptions, it is assumed that the noise components are uncorrelated, i.e. that f_u (or in the static case Σ_u) is diagonal. In other words the basic idea is not to look for the best approximation of x_t by the latent variables χ_t as in the case of PCA, but to separate the common components described by the factors from the individual components described by "noise". The factors here have a splitting property: For given factors, the components of (x_t) are conditionally uncorrelated. The static factor

model with idiosyncratic noise is the classical factor model, with a long history dating back, as has been mentioned above, to the beginning of the twentieth century.

Commencing from given Σ_x, we see from (4), that in the static case the following two identifiability problems arise:

1. For given Σ_x, what is the set of all pairs $\Sigma_\chi = \Lambda\Sigma_\xi\Lambda'$ and Σ_u, where Σ_χ is positive semidefinite, singular and symmetric and Σ_u is positive semidefinite and diagonal, such that (4) is satisfied? In this set r, i.e. the rank of Σ_χ, may vary; we restrict ourselves to the subset, where r is minimal and from now on the letter r is used for such a minimal r.
2. What is the set of all Λ and Σ_ξ corresponding to Σ_χ? In most cases, $\Sigma_\xi = I_r$ is assumed, so that Λ is unique up to right-multiplication by orthogonal matrices (factor rotation).

As far as the first problem is concerned, the answer is, that, in general, Σ_χ and Σ_u are not uniquely defined for given Σ_x, see [16], but they are generically unique if r is smaller than or equal to the so called Lederman bound $\frac{2n+1}{2} - \sqrt{\frac{(2n+1)^2}{4} - n^2 + n}$. (The Lederman bound follows from counting the numbers of free parameters on both sides of (4)).

For the case that r is smaller or equal to the Lederman bound, estimators of Λ and Σ_u are obtained from maximizing the Gaussian log-likelihood function under suitable normalization conditions on Λ guaranteeing uniqueness, and the ML-estimators can be shown to be consistent under general assumptions, see [1]. In the quasi static case this function is no longer the log-likelihood, but nevertheless yields consistent estimators.

In contrast to the PCA model, here r, or, to be more precise, the minimal r in all decompositions (4) is intrinsic, i.e. it is a property of Σ_x. Tests for determining r have been proposed in [2]. Furthermore for the factor model with idiosyncratic noise, the factors, in general, are not functions of the observations and thus have to be approximated by (linear) functions of the observations.

For dynamic factor models with idiosyncratic noise, a rather complete structure theory has been developed in [18]. The answer to the dynamic analogon of the first identifiability question above is, that, for given spectral density f_x, the spectra $f_\chi = \Lambda(e^{-i\lambda})f_\xi(\lambda)\Lambda(e^{-i\lambda})^*$ and f_u are generically unique for $r \leq n - \sqrt{n}$.

As has been shown in [18], for dynamic factor models with idiosyncratic noise the set of all spectral densities $f_x(\lambda)$ described by (3) for given r is a "thin" subset of the set of all spectral densities $f_x(\lambda)$, if $r < n - \sqrt{n}$ holds.

For estimation and specification in the dynamic case we refer to [10], [15] and

[23].

In this area there is still a substantial number of unsolved problems.

4. Generalized Factor Models. The classical assumption that f_u (or in the static case Σ_u) is diagonal turns out to be too restrictive for many applications, where e.g. "local" dependency between the noise components may occur. Moreover, in a number of applications, e.g. in cross-country business cycle analysis, asset pricing [7] or in monitoring and forecasting economic activity by estimation of common factors ("diffusion indexes") [20], the cross-sectional dimension may be high, possibly exceeding sample size.

Both, the issue of weakening the assumption of uncorrelatedness of the idiosyncratic components and the demand for modelling high dimensional time series has lead to the development of generalized factor models, see for example [7], [11], [12], [13], [9].

For the corresponding analysis the cross-sectional dimension n is not regarded as fixed. Thus we consider a double sequence $(x_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$, where our general assumptions hold true for every vector $x_t^n = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ with $n \in \mathbb{N}$. Hence, using an obvious notation, we have a sequence of factor models

$$x_t^n = \Lambda^n(z)\xi_t + u_t^n,\ t \in \mathbb{Z},\ n \in \mathbb{N}, \tag{10}$$

where the noise vector u_t^n and the transfer functions $\Lambda^n(z)$ are nested (in the sense that e.g. the coefficient matrices Λ_j^m are the $m \times r$ top submatrices of Λ_j^n for $m \leq n$ and for all $j \in \mathbb{Z}$); the r-dimensional factors ξ_t do not depend on n.

In order to render explicit the dependence on n, the previously defined symbols may be provided with a superscript n, e.g. f_u^n denotes the spectral density of (u_t^n).

Weak dependence between the noise components in u_t^n (substituting the classical assumption of uncorrelatedness) is formalized by requiring the largest eigenvalue of the spectral density f_u^n, $\omega_{u,1}^n : [-\pi, \pi] \to \mathbb{R}$ say, to be uniformly bounded for all $n \in \mathbb{N}$, i.e.

ASSUMPTION 4.1. *There exists a $\overline{\omega} \in \mathbb{R}$, such that $\omega_{u,1}^n(\lambda) \leq \overline{\omega}$ for all $\lambda \in [-\pi, \pi]$ and for all $n \in \mathbb{N}$.*

On the other hand, the next assumption guarantees that every factor has a minimum amount of influence on infinitely many observations:

ASSUMPTION 4.2. *The first r eigenvalues of f_χ^n, $\omega_{\chi,j}^n$ say, $j = 1, \ldots, r$ diverge a.e. in $[-\pi, \pi]$ as n tends to infinity.*

Equation (10) together with the assumptions imposed so far describe the gener-

alized (dynamic) factor model (GFM and GDFM respectively).

A basic idea in dealing with GFMs is to obtain an increasing amount of information from adding time series by averaging out the noise term. As a simple example consider the following one-factor GFM with $\Lambda^n = (1, 1, \ldots, 1)'$, ξ_t i.i.d. (i.e. independent and identically distributed) and u_{it} i.i.d. in i and t. Then, as $\frac{1}{n}\sum_i u_{it}$ converges to 0, the noise term is averaged-out.

This example is used to demonstrate the following general result: As shown in [11], assumption 4.1 implies, that $k^n(z)u_t^n$ converges to 0 in mean square for every sequence of transfer functions $k^n(z) = \sum_{j=-\infty}^{\infty} k_j^n z^j$, $k_j^n \in \mathbb{R}^{1\times n}$ with $\|k^n\| \to 0$ for $n \to \infty$. Here we use the norm defined by $\|k^n\|^2 = \int k^n(e^{-i\lambda})k^n(e^{-i\lambda})^* d\lambda$.

Forni and Lippi, see [11], provide necessary and sufficient conditions for the existence of an underlying GFM in terms of the observable spectral densities $f_x^n, n \in \mathbb{N}$:

The double sequence $(x_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ can be represented by a sequence of GFMs, if and only if,

1. the first r eigenvalues of f_x^n, $\omega_{x,j}^n$ say, $j = 1, \ldots, r$, diverge a.e. in $[-\pi, \pi]$ as n tends to infinity,
2. the $(r+1)$-th eigenvalue of f_x^n, $\omega_{x,r+1}^n$ say, is uniformly bounded for $\lambda \in [-\pi, \pi]$ a.e. and for all $n \in \mathbb{N}$.

As will be discussed below (admittedly in a sloppy manner), the sequence of PCA-models for $(x_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ will converge to the sequence of corresponding GDFMs, and the spectral densities of χ_t^n and u_t^n of the GDFM are "asymptotically (for $n \to \infty$) identifiable" by the spectral densities corresponding to dynamic PCA obtained from the canonical representation of f_x^n,

$$f_x^n(\lambda) = O_1^n(e^{-i\lambda})\Omega_1^n(\lambda)O_1^n(e^{-i\lambda})^* + O_2^n(e^{-i\lambda})\Omega_2^n(\lambda)O_2^n(e^{-i\lambda})^*, \tag{11}$$

where Ω_1^n, Ω_2^n, O_1^n and O_2^n are defined as in (8).

As shown in Forni et al. [12] the sequence of PCA-models for $(x_t^n), n \in \mathbb{N}$, approximates the corresponding sequence of GDFMs in the following sense:

- The Hilbert space spanned by the first r dynamic principal components of (x_t^n), i.e. the space spanned by the scalar components of $O_1^n(z)^* x_t^n, t \in \mathbb{Z}$, in the Hilbert space of all square-integrable (scalar) random variables, converges to the space spanned by the scalar components of (ξ_t), where convergence of spaces is understood in the sense that the perpendiculars of a projection from one space to the other converge to 0 in mean square.
- Every component of the latent variables (or the noise resp.) from the PCA-model converges to the corresponding generalized factor model variable as n

tends to ∞. (Note that here convergence of one-dimensional random variables is in the sense of mean squares convergence.) In other words, let $\chi_{it}^n = o_{1,i}^n(z)O_1^n(z)^* x_t^n$ denote the projection of x_{it}^n onto the space spanned by the one-dimensional components of $O_1^n(z)^* x_t^n$, $t \in \mathbb{Z}$, i.e. the i-th element of the latent variable of the corresponding PCA-model at time t, then

$$\lim_{n\to\infty} \chi_{it}^n = \chi_{it}, \text{ for all } i \in \mathbb{N} \text{ and for all } t \in \mathbb{Z}, \tag{12}$$

where χ_{it} denotes the i-th element of $\Lambda^n(z)\xi_t$ (for $n \geq i$), hence the corresponding "true" latent variable.

Thus, the dynamic PCA-model and the generalized dynamic factor model are asymptotically equivalent, in the above sense.

Concerning identifiability (which sloppy speaking relates to an "infinite sample size T"), the results from dynamic PCA carry over, i.e. asymptotically, as n tends to infinity, the latent variables as well as the idiosyncratic components are identifiable, whereas even for $n \to \infty$ the transfer functions $\Lambda^n(z)$ and the factors ξ_t are only identifiable up to regular dynamic linear transformations $\sum_{j=-\infty}^{\infty} T_j z^j$. If the factors are assumed to be white noise with $\mathrm{E}\xi_t\xi_t' = I_r$, which is no restriction on f_χ, since our assumptions always allow this transformation, the factors are identifiable up to "static rotations", i.e. constant orthogonal matrices.

Let us continue our simple example from above. If we assume that $\mathrm{var}(\xi_t) = 1$ and $\Sigma_u^n = I_n$, then $\Sigma_x^n = \mathbf{1}_n + I_n$, where $\mathbf{1}_n$ denotes the $n \times n$ matrix consisting of ones. The eigenvalue decomposition of Σ_x^n yields $\omega_{x,1}^n = n+1$ as the largest eigenvalue and the corresponding eigenvector is $O_1^n = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)'$, so that the variance matrix of the latent variables of the PCA model equals $O_1^n(n+1)O_1^{n'}$, whose elements converge to the elements of the corresponding GFM-matrix, i.e. $\Sigma_\chi^n = O_1^n n O_1^{n'}$, as $n \longrightarrow \infty$.

For estimation, the dynamic PCA estimators corresponding to (8) and (9), as discussed in Section 2, may be employed, thus e.g. $\hat{\chi}_t^n = \left[\hat{O}_1^n(z)\hat{O}_1^n(z)^*\right]_t x_t^n$, where $\hat{O}_1^n(e^{-i\lambda})$ denotes the matrix consisting of the first r eigenvectors of a consistent estimator of $f_x^n(\lambda)$, $\hat{f}_x^n(\lambda)$ say, and the subscript t indicates that the filter $\hat{O}_1^n(z)\hat{O}_1^n(z)^*$, that is in general two-sided and of infinite order, has to be truncated at lag $t-1$ and lead $T-t$, as for $t \leq 0$ and $t > T$, x_t^n is not available. For fixed n and $T \to \infty$ a consistent estimator of f_x^n yields consistent estimators of the spectral densities $\check{f}_\chi^n$ and $\check{f}_u^n$ respectively of the latent variables and the noise resp. of the corresponding PCA-model, since eigenvalues and suitably normalized eigenvectors are continuous functions of the original frequency-dependent matrices. As a consequence of this consistency result the estimators of the latent variables in the PCA model converge to the true values of the PCA model and the same holds true for the noise term. However,

as a consequence of the truncation of $\hat{O}_1^n(z)\hat{O}_1^n(z)^*$ this convergence of the estimators of χ_t^n and u_t^n, as $T \to \infty$, can only be granted for a "central" part of the sample, whereas for fixed t the estimators may never be consistent. An analogous statement holds for the latent variables and the noise respectively of the GFM as n and T tend to ∞.

So far the number of factors r was considered fixed and known, whereas in practice it has to be determined from the data. The above discussion indicates that the eigenvalues of $\hat{f}_x^n$ could be used for determining the number of factors and Forni et al. in [12] propose a heuristic rule, but indeed, no formal testing procedure has been developed yet. For the static case an information criterion for estimating the number of factors has been proposed in [3].

While the two-sidedness of the filters occurring in dynamic PCA already causes problems for estimation at the edges of the observations, the fact that the filters are in general two-sided and thus non causal may yield "infeasible naive" forecasts for $\chi_{t+h}^n, h > 0$, based on $\chi_s^n, s \leq t$, since in general the latent PCA-variable χ_t^n may depend on $x_s^n,\ s > t$. One way to overcome this difficulty, see([13]), is to assume that $\Lambda^n(z)$ is polynomial of degree p, i.e.

ASSUMPTION 4.3. *$\Lambda^n(z)$ is of the form $\Lambda^n(z) = \sum_{j=0}^p \Lambda_j^n z^j$,*

and to restrict the factors to stationary AR-processes of finite order $s \leq p+1$:

ASSUMPTION 4.4. *(ξ_t) is of the form $\xi_t = A(z)^{-1}\varepsilon_t$, $A(z) = I - A_1 z - \ldots - A_s z^s$ with $A_j \in \mathbb{R}^{r\times r}$, $\det A(z) \neq 0$ for $|z| \leq 1$, $s \leq p+1$ and the innovations ε_t are r-dimensional white noise with $\mathrm{E}\varepsilon_t\varepsilon_t' = \Sigma_\varepsilon > 0$ (and are uncorrelated with u_t^n at any leads and lags).*

The model then can be written in quasi static form, on the cost of higher dimensional factors, as:

$$x_t^n = \bar{\Lambda}^n F_t + u_t^n = \chi_t^n + u_t^n,\ t \in \mathbb{Z},\ n \in \mathbb{N}, \tag{13}$$

where $F_t = (\xi_t', \ldots, \xi_{t-p}')'$ is the $q = r(p+1)$-dimensional vector of stacked factors and $\bar{\Lambda}^n = (\Lambda_0^n, \ldots, \Lambda_p^n)$ is the $(n \times q)$-dimensional static factor loading matrix. Note that $(x_t^n), (\chi_t^n), (u_t^n)$ and their variances and spectra as well as $\bar{\Lambda}^n$ still depend on n, but for sake of legibility we will drop the superscript from now. Under the assumptions imposed, F_t and u_t remain uncorrelated at any leads and lags and thus f_x is of the form

$$f_x(\lambda) = \bar{\Lambda} f_F(\lambda)\bar{\Lambda}^* + f_u(\lambda). \tag{14}$$

For estimation, Stock and Watson [20] propose the quasi-static PCA procedure with q factors and they prove consistency for the factor estimates (i.e. the first q

sample principal components of x_t) up to premultiplication with a nonsingular matrix as n and T tend to infinity under general assumptions. Sloppy speaking, the space spanned by the true GFM-factors can be consistently estimated. As it is easily seen, this is just a special case of the dynamic results discussed above.

An alternative two-stage "generalized PCA" estimation method has been proposed by Forni et al. in [13]. It differs from classical PCA in two respects: First in the estimation of the covariances of the common components of the PCA and second in the determination of the weighting scheme. While classical PCA is based on the sample covariance $\hat{\Sigma}_x$ of (x_t), the approach in [13] commences from the estimated spectral density $\hat{f}_x$ decomposed according to dynamic PCA, see (11). Then the covariance matrices Σ_χ and Σ_u of the common component and the noise respectively are estimated as

$$\hat{\Sigma}_\chi = \int_\pi^{-\pi} \hat{f}_\chi(\lambda) d\lambda \qquad \text{and} \qquad \hat{\Sigma}_u = \int_\pi^{-\pi} \hat{f}_u(\lambda) d\lambda, \tag{15}$$

or to be more precise, since $\hat{f}_x$ is defined on a finite grid of frequencies, by the corresponding sums. In the second step, the factors are estimated as $C'x_t$, where the $(n \times q)$-dimensional weighting matrix C consists of the first q generalized eigenvectors of the matrices $(\hat{\Sigma}_\chi, \hat{\Sigma}_u)$, i.e. $\hat{\Sigma}_\chi C = \hat{\Sigma}_u C \Omega_1$, where Ω_1 denotes the diagonal matrix containing the q largest generalized eigenvalues. C is thus the matrix that maximizes

$$\begin{aligned} &\text{tr } (C'\hat{\Sigma}_\chi C) \\ &\text{s.t. } C'\hat{\Sigma}_u C = I_q, \end{aligned} \tag{16}$$

(whereas in static PCA the corresponding weights C maximize tr $(C'\hat{\Sigma}_x C)$, s.t. $C'C = I_q$). The matrix Λ is then estimated by projecting x_t onto $C'x_t$, i.e.

$$\hat{\Lambda} = \hat{\Sigma}_x C (C'\hat{\Sigma}_x C)^{-1}.$$

The common component estimator $\hat{\chi}_t$ defined this way can be shown to be consistent (for n and T tending to infinity). The argument for this procedure is that less weight is given to variables with higher noise variance and lower common component variance respectively and vice versa, thus the common-to-idiosyncratic variance ratio in the resulting latent variables is maximized, which could possibly improve efficiency upon static PCA.

Boivin and Ng [4] propose a third estimation method called "weighted PCA" which is related to classical PCA in the same way as generalized least squares (GLS) is related to ordinary least squares (OLS) in the regression context. Remember, that, if the regression residuals are non-spherical, but the structure of the variance matrix is known, GLS weights the residuals proportional to the inverse of the square

root of their variance matrix and yields (linearly) efficient estimators. Assume for a moment that the noise variance Σ_u were known, then the same principle could be applied to PCA by transforming the least squares problem (5), i.e. $\min \mathrm{E} u_t' u_t$, into the generalized least squares problem

$$\min_{\Lambda, C} \mathrm{E} u_t' \Sigma_u^{-1} u_t = \min_{\Lambda, C} \mathrm{E}\big(x_t - \Lambda C x_t\big)' \Sigma_u^{-1} \big(x_t - \Lambda C x_t\big), \tag{17}$$

which is solved for

$$\Lambda = \Sigma_u^{1/2} O_1 \text{ and } C = O_1' \Sigma_u^{-1/2}, \tag{18}$$

where now O_1 denotes the first q eigenvectors of $\Sigma_u^{-1/2} \Sigma_x \Sigma_u^{-1/2}$. For estimation, Σ_x is replaced by the corresponding sample covariance. Additionally an estimate of Σ_u is needed. Since the residual matrix of the unweighted PCA-model is singular, it cannot be used. A feasible alternative is the estimator resulting from dynamic PCA, i.e.

$$\hat{\Sigma}_u = \int_{\pi}^{-\pi} \hat{f}_u(\lambda) d\lambda, \tag{19}$$

which in many cases, but not always, will be nonsingular. An alternative estimator is proposed in [4]. Again the factor space is consistently estimated as n and T tend to infinity.

Whether the estimation methods described above can improve upon conventional static PCA, is still a partially open question. At least from a theoretical point of view, the question of (asymptotic) efficiency gain is not yet answered. In [13] and [4] Monte Carlo experiments are described showing that the last two estimators described above are better than static PCA in certain cases, but the results are highly dependent on the properties of the underlying stochastic processes. On the other hand the results of an empirical study in the context of macroeconomic forecasting [21] do not show much difference.

Concerning forecasting in the quasi-static model, at least two different approaches have been proposed. First, as suggested by Forni et al. in [13], prediction can be restricted to the common component. In this case the variable to be forecast, $x_{i,t+h}$ say, is projected onto the estimate of $\chi_{i,t}$ or equivalently F_t. Second, as proposed by Stock and Watson in [21], the information contained in the idiosyncratic component $(u_{i,t})_{t \in \mathbb{Z}}$ can be additionally taken into account, in which case $u_{i,t}$ is supposed to follow the AR(S)-process $b_i(z) u_{i,t} = \nu_{i,t}$ with $\nu_{i,t}$ being white noise. Let us consider one-step ahead forecasts. Under these assumptions the factor model equation for $x_{i,t+1}$, i.e.

$$x_{i,t+1} = \lambda_i F_{t+1} + u_{i,t+1}, \tag{20}$$

where λ_i denotes the i-th row of Λ, can be rewritten as

$$x_{i,t+1} = b_i(z)\lambda_i F_{t+1} + \gamma_i(z)x_{i,t} + \nu_{i,t+1}, \tag{21}$$

where $\gamma_i(z) = (1 - b_i(z))z^{-1}$. Thus the one-step ahead forecast, $\hat{x}_{i,t,1}$ say, is given by

$$\hat{x}_{i,t,1} = \beta_i(z)F_t + \gamma_i(z)x_{i,t}, \tag{22}$$

where $\beta_i(z)F_t$ is the projection of $b_i(z)\lambda_i F_{t+1}$ onto the space spanned by present and past values of the one-dimensional components of (F_t). The forecast is thus computed via a "factor augmented AR model".

So far we have discussed several methods to estimate the factor space, but we were not concerned with the identification of either the factors F_t or the factor loading matrix Λ. Note, that although e.g. the first q static principal components (asymptotically for $n \longrightarrow \infty$) form a basis of the factor space and thus their estimates can be used as factor estimates as long as one is interested in projections onto that space e.g. for estimating or forecasting the latent variables. This choice is arbitrary since any nonsingular linear transformation of those estimates span the same space. We are now concerned with the identification of a linear transformation that allows for a meaningful "structural" interpretation of the resulting factors and factor loading matrix, see Forni et al. [14].

Having estimated the dynamic model in its static form (13), to identify the r-dimensional dynamic factors ξ_t (or to be more precise the space spanned by those), it is of particular interest to reveal the structure of F_t, i.e. the fact that the stacked factors F_t follow a structured AR(1)-process,

$$F_t = DF_{t-1} + e_t, \tag{23}$$

where

$$D = \begin{pmatrix} A_1 & A_2 & \dots & A_s & 0 \\ (r \times r) & & \dots & (r \times r) & (r \times (r(p+1-s))) \\ I & & & & 0 \\ (pr \times pr) & & & & (pr \times r) \end{pmatrix}$$

and $e_t = (\varepsilon_t', 0, \dots, 0)'$ is orthogonal to F_{t-1}.

In general this property will not be kept by any of the factor estimates discussed above. Let now denote by G_t the population analogon of a factor estimate according to one of the methods discussed above (e.g. the first q static principal components

for infinite cross-section if the population variance Σ_x were known), then $G_t = HF_t$ is satisfied with H nonsingular. As a consequence G_t has the AR(1)-representation

$$G_t = HDH^{-1}G_{t-1} + He_t. \tag{24}$$

We observe that $HDH^{-1} = \mathrm{E}(G_tG'_{t-1})\Sigma_G^{-1}$ and the residuals He_t can be written as $He_t = H_r\varepsilon_t = (H_rP^{-1})P\varepsilon_t = O_1\Omega_1^{1/2}P\varepsilon_t$, where H_r denotes the first r columns of H, P is a regular $r \times r$ matrix, Ω_1 and O_1 result from a PCA for the noise term on the r.h.s. in (24), thus Ω_1 is the diagonal matrix containing the r largest (i.e. non-zero) eigenvalues of $H_r\Sigma_\varepsilon H'_r$ and O_1 is the matrix of corresponding eigenvectors. Hence we have the solution

$$G_t = (I - HDH^{-1}z)^{-1}O_1\Omega_1^{1/2}P\varepsilon_t. \tag{25}$$

Recalling (13), we have th at, $\chi_t = \Lambda(z)\xi_t = \Lambda F_t = \tilde{\Lambda}G_t$ with $\tilde{\Lambda} = \Lambda H^{-1} = E(x_tG'_t)\Sigma_G^{-1}$, and thus χ_t can be written as

$$\chi_t = \tilde{\Lambda}(I - HDH^{-1}z)^{-1}O_1\Omega_1^{1/2}P\varepsilon_t. \tag{26}$$

Estimators for the transfer function $\Lambda(z)A(z)^{-1}$ and the r-dimensional dynamic factor innovations ε_t are derived in an analogous way by replacing G_t by an estimator discussed above and the population second moments by their sample counterparts. The so defined estimators can be shown to be consistent (as n and $T \longrightarrow \infty$), see [14].

As a consequence the transfer function $\Lambda(z)A(z)^{-1}$ and the r-dimensional dynamic factor innovations ε_t are identifiable up to static linear regular transformations. This last identification problem may be solved by choosing P in a way such that $\Lambda(z)A(z)^{-1}P$ has a meaningful interpretation or by imposing zero restrictions on e.g. $\Lambda(0)A(0)^{-1}P$.

REFERENCES

[1] T. W. Anderson, *Estimating Linear Statistical Relationships*, Annals of Statistics, 12(1984), pp. 1-45.

[2] T. W. Anderson and H. Rubin, *Statistical Inference in Factor Analysis*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, V:111-150, 1956

[3] J. Bai and S. Ng, *Determining the Number of Factors in Approximate Factor Models*, Econometrica, 70:1(2002), pp. 191-221.

[4] J. Boivin and S. Ng, *Are More Data Always Better for Factor Analysis?* J. of Econometrics, 2004

[5] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden Day, San Francisco, 1981

[6] C. Burt, *Experimental tests of general intelligence*, British Journal of Psychology 3(1909), pp. 94-177.

[7] G. Chamberlain and M. Rothschild, *Arbitrage, Factor Structure and Meanvariance Analysis on Large Asset Markets*, Econometrica, 51:5(1983), pp. 1281-1304.

[8] M. Deistler and E. Hamann, *Identification of Factor Models for Forecasting Returns*, J. of Financial Econometrics, 2005.

[9] M. Deistler, *Linear Models for Multivariate Time Series*, M. Winterthaler et al., eds., Handbook of Time Series Analysis, Wiley-VCH, 2006.

[10] R. F. Engle and M. F. Watson, *A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates*, J. of the American Statistical Association, 76(1981), pp. 774-781.

[11] M. Forni and M. Lippi, *The Generalized Dynamic Factor Model: Representation Theory*, Econometric Theory, 17(2001), pp. 1113-1141.

[12] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The Generalized Dynamic Factor Model: Identification and Estimation*, The Review of Economic Studies, 65(2000), pp. 453-473.

[13] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, *The Generalized Dynamic Factor Model: One-sided Estimation and Forecasting*, Journal of the American Statistical Association, 100(2005), pp. 830-840.

[14] M. Forni, D. Giannone M. Lippi, and L. Reichlin, *Opening the Black Box: Structural Factor Models versus Structural VARs*, Manuscript, Universite Libre de Bruxelles, 2005.

[15] J. Geweke, *The Dynamic Factor Analysis of Economic Time Series Models*, chapter 19, 365-383, North-Holland Amsterdam, 1977.

[16] D. N. Lawley and A. E. Maxwell, *Factor Analysis as a Statistical Method*, Butterwoth & Co., 2nd ed., 1971.

[17] T. J. Sargent and C. A. Sims, *Business Cycle Modelling without Pretending to have too much a Priori Economic Theory*, New Methods in Business Cycle Research, Federal Reserve Bank of Minneapolis, 1977.

[18] W. Scherrer and M. Deistler, *A Structure Theory for Linear Dynamic Errors-in-variables Models*, SIAM, J. on Control Optim., 36:6(1998), pp. 2148-2175.

[19] C Spearman, *General Intelligence, Objectively Determined and Measured*, American J. of Psychology, 15(1904), pp. 201-203.

[20] J. H. Stock and M. W. Watson, *Diffusion Indexes*, NBER Working Paper 6702, 1998

[21] J. H. Stock and M. W. Watson, *An Empirical Comparison of Methods for Forecasting Using Many Predictors*, Mimeo, Princeton University, 2005

[22] L. L. Thurstone, *The vectors of the mind*, Psychological Review, 41(1934), pp. 1-32.

[23] M. W. Watson and R. F. Engle, *Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models*, J. of Econometrics, 23(1983), pp. 385-400.