Wiley StatsRef:
Statistics Reference Online

Factor Analysis, Dynamic

By *Marianna Bolla*

Keywords: *standard factor analysis, dynamic factor analysis, generalized dynamic factor models, asymptotic inference using principal components, parametric view*

Abstract: We review small- and medium-sized standard and dynamic factor analysis and include recent asymptotic results on high-dimensional dynamic factor models. The latter allow the construction of consistent estimators in the factor model, which are based on principal component analysis techniques either in the time or in the frequency domain. The point is that when both the number of time series and the number of observations (time instances) tend to infinity, the error terms become negligible and the spectral gap in the spectral density decides the number of hidden factors. Although mainly developed within Time Series Econometrics, high-dimensional factor models are quite appealing for applications to environmental data sets and for expanding (both in time and in the number of observables) networks. Eventually, in the small word of real-life data (in the presence of sudden economic changes and crises, macroeconomic time series span sometimes for 10–20 years only), a parametric model with an easily implementable algorithm and a numerical example is presented.

1 Introduction

02443

In **factor analysis**, both standard and dynamic, the variability of a set of observed stochastic variables is accounted for by a small (relative to the data set dimension) number of unobservable *common factors*, plus variable-specific causes of variation. Factor models belong, therefore, to the wide class of techniques used to reduce the dimension, hence overcoming the "curse of dimensionality" of observed data sets. In particular, dynamic factor models can be viewed as representations of high-dimensional **stochastic processes** that are parsimonious and, therefore, well suited for prediction (*see* **Forecasting, Environmental**). On the other hand, depending on the data set under consideration, the factors, after adequate linear transformation, can be related to causes of variation such as sources of pollution in atmospheric studies, demand and supply shocks in macroeconomic applications, market confidence and external influences in stock prices, and so on.

03025
07702

Budapest University of Technology and Economics, Budapest, Hungary

Update based on original article by Marco Lippi, Wiley StatsRef: Statistics Reference Online © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

Dynamic factor analysis (FA) is concerned with a family of stochastic processes, which can be represented as a *common component*, driven by the factors plus a *process-specific*, or *idiosyncratic component*. Typically, the factors and the specific components are unobserved mutually orthogonal stochastic processes, and the specific components belonging to different processes are orthogonal to each other. Both assumptions, with some variations, are crucial for identification and estimation of common and specific components.

In Section 2, the basic notions relative to standard factor models, principal component factor, and the recently developing independent component analysis (ICA) are recalled. Section 3 is dedicated to dynamic FA, including the nonparametric setup, the recent studies on high-dimensional dynamic factor models, and techniques allowing estimation of the factor structure that are alternatives to the standard ones. It is seen that **principal components**, which is a valid tool only in special cases within small- or medium-sized factor models, provide consistent estimators for common and specific components when the number of observations and the number of variables tend to infinity. A parametric model using finite lags and applicable for prediction in moderate size time series is also introduced and illustrated via a numerical example.

Although high-dimensional factor models have been mainly developed within Time Series Econometrics (*see* **Econometrics**) with applications in macroeconomic analysis, they are perfectly suitable for applications to financial and environmental data as well.

2 Standard Factor Analysis Methods for Reduction of Data Dimension

2.1 Principal Component Analysis (PCA)

Consider an n-dimensional *random vector* (*rv*) $\mathbf{X} = (X_1, \dots, X_n)'$ (the vectors are columns, and $'$ denotes the transposition). Let $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$ denote its *expectation (vector)* and $\boldsymbol{C} = \mathrm{Var}(\mathbf{X}) = \mathbb{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$ its covariance matrix. In many practical applications, $\mathbf{X}$ has n-dimensional normal (Gaussian) distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{C}$, denoted by $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{C})$. We usually assume that the distribution of $\mathbf{X}$ is not degenerate, that is, rank $\boldsymbol{C} = n$.

Performing principal component analysis (PCA) on $\mathbf{X}$ means the transformation

$$\mathbf{X} = \boldsymbol{U}\mathbf{Y} + \boldsymbol{\mu} \tag{1}$$

where the n dimensional rv $\mathbf{Y}$ of the *principal components* (*PC*s) has zero expectation and uncorrelated components, with diagonal covariance matrix $\boldsymbol{\Lambda}$. Note that $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are obtained from the **Spectral Decomposition** $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$, where the diagonal of $\boldsymbol{\Lambda}$ contains the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ of $\boldsymbol{C}$, whereas the columns $\mathbf{u}_1, \dots, \mathbf{u}_n$ of $\boldsymbol{U}$ are the corresponding orthonormal eigenvectors. As $\boldsymbol{U}^{-1} = \boldsymbol{U}'$, the PCs are obtained by the transformations

$$Y_i = \mathbf{u}_i'(\mathbf{X} - \boldsymbol{\mu}), \quad \text{where} \quad \mathrm{Var}(Y_i) = \lambda_i \quad (i = 1, \dots, n) \tag{2}$$

As $\sum_{i=1}^n \lambda_i = \mathrm{tr}\ \boldsymbol{C} = \sum_{i=1}^n \mathrm{Var}(X_i)$, in this order, the PCs explain the largest possible part of the total variation of $\mathbf{X}$, characterized as follows.

Theorem 1. *The variance of Y_1 of Equation (2) is the largest possible among the variances of linear combinations $\mathbf{v}'(\mathbf{X} - \boldsymbol{\mu})$, subject to $\|\mathbf{v}\| = 1$. Successively, the variance of Y_k of Equation (2) is the largest possible among the variances of linear combinations $\mathbf{v}'(\mathbf{X} - \boldsymbol{\mu})$ that are uncorrelated with $Y_1, \dots, Y_{k-1}$, subject to $\|\mathbf{v}\| = 1$, for $k = 2, \dots, n$.*

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

A bit more is true. For every $1 \le k \le n$, the first k PCs give the best k-dimensional approximation of $\mathbf{X}$ in the following sense: the minimum of $\mathbb{E} \parallel \mathbf{X} - \boldsymbol{P}\mathbf{X} \parallel$ with an $n \times n$ matrix $\boldsymbol{P}$ of rank k is attained by the projection $\boldsymbol{P}^*$ onto the k-dimensional subspace spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_k$. In this new coordinate system, $\boldsymbol{P}^*\mathbf{X} = (Y_1, \ldots, Y_k, 0, \ldots, 0)'$.

We often look for the smallest possible number k of PCs that explain a relatively large part of the total variance of $\mathbf{X}$. One can consider the successive ratios $\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}$ and select a k such that there is a "gap" in the spectrum of $\boldsymbol{C}$ between λ_k and λ_{k+1}. When we have an N-element sample ($N > n$) from a multivariate normal distribution (in which case the PCs are not only uncorrelated but also independent), we can perform a likelihood ratio test for testing the following sequence of null hypotheses:

$$H_{0,k} : \lambda_{k+1} = \cdots = \lambda_n \quad \text{for} \quad k = 0, 1, \ldots, n-1$$

until accepted. By the asymptotic theory of the likelihood ratio tests, the transformed test statistic $-2 \ln T_{n,k}$ has the form

$$N(n-k) \ln \frac{a}{g} \quad \text{with} \quad a = \frac{\hat{\lambda}_{k+1} + \cdots + \hat{\lambda}_p}{p-k}, \quad g = (\hat{\lambda}_{k+1} \ \ldots \ \hat{\lambda}_p)^{\frac{1}{p-k}}$$

where $\hat{\lambda}_i$s are the eigenvalues of the empirical covariance matrix $\hat{\boldsymbol{C}}$ of the sample, and for "large" N, it asymptotically follows χ^2-distribution with degrees of freedom $\frac{1}{2}(n-k+2)(n-k-1)$ (the decrease in the number of parameters under the assumption of $H_{0,k}$). Given the significance, we stop if $H_{0,k}$ is accepted, which can be interpreted as the number of significant PCs is k. The PCs themselves are estimated from the sample via its mean vector and the spectral decomposition of $\hat{\boldsymbol{C}}$.

2.2 Factor Analysis (FA)

In this article, a smaller number of latent variables explain the correlations between the original ones. We say that the n-dimensional rv $\mathbf{X}$ has a factor structure if each variable X_i depends on a small number of latent common factors plus a component that is specific to X_i. Formally, $\mathbf{X}$ has a k-factor structure if it obeys the following model with the integer $1 \le k < n$:

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{B}\mathbf{f} + \mathbf{e} \tag{3}$$

where the components of the k-dimensional rv $\mathbf{f} = (f_1, \ldots f_k)'$ are the *common factors* and the components of the n-dimensional rv $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)'$ are the *individual factors* (disturbances), whereas the $n \times k$ matrix $\boldsymbol{B} = (b_{ij})$ contains the *factor loadings*. We make the following assumptions:

$$\mathbb{E}(\mathbf{f}) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{f}) = \boldsymbol{I}_k, \quad \mathbb{E}(\mathbf{e}) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{e}) = \boldsymbol{D}, \quad \mathrm{Cov}(\mathbf{f}, \mathbf{e}) = \boldsymbol{O} \tag{4}$$

where $\boldsymbol{D}$ is an $n \times n$ diagonal matrix and the cross-covariance matrix of $\mathbf{f}$ and $\mathbf{e}$, denoted by $\mathrm{Cov}(\mathbf{f}, \mathbf{e})$, is the $k \times n$ zero matrix. This means that both the common and the individual factors have uncorrelated components that are also uncorrelated with each other; further, the factors are normalized so that they have unit variances. If $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{C})$, then $\mathbf{Y} \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{I}_k)$ is a k-dimensional standard normal rv. However, its components cannot be obtained with an explicit transformation, like the PCs. The factors are latent variables that we cannot observe directly, we can only estimate the so-called factor scores.

To identify the model (3), consider the equation

$$\boldsymbol{C} = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{D} \tag{5}$$

DOI: 10.1002/9781118445112.stat07505.pub2

04464

obtained by equating the covariance matrices. This equation is the basis for the **maximum likelihood** (ML) estimation of the rank k matrix $\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}$ and the diagonal matrix $\boldsymbol{D}$; further, for testing the hypothesis that the number of factors is k. For the coordinates and variances of X_is, Equations (3) and (5) provide

$$X_i = \mu_i + \sum_{j=1}^{k} b_{ij} f_j + e_i, \quad \mathrm{Var}(X_i) = \sum_{j=1}^{k} b_{ij}^2 + \mathrm{Var}(e_i), \quad i = 1, \ldots, n$$

That is, every X_i depends on all of the common factors f_js, but only depends on its own individual factor e_i. Here, $\sum_{j=1}^{k} b_{ij}^2$ is the part of the variance of X_i, accounted for the common factors, and it is called *communality* of X_i; this makes sense when, instead of $\boldsymbol{C}$, the correlation matrix of X_is is used (it indeed has rational if X_is are measured on different scales).

In Ref. 1, via counting the number of parameters, it is proved that unique solution to Equation (5) can be expected with the so-called Lederman bound $k \leq \frac{1}{2}(2n + 1 - \sqrt{8n+1})$. Also observe that the structure described in Equations (3) and (4) is not sufficient to identify the factors and the factor loadings: if $\boldsymbol{Q}$ is a $k \times k$ orthogonal matrix, then $\boldsymbol{Q}\mathbf{f}$ and $\boldsymbol{B}\boldsymbol{Q}^{-1}$ fulfill Equations (3) and (4) as well as $\mathbf{f}$ and $\boldsymbol{B}$ do. However, when the factors and factor loadings are linearly transformed as above, the common components $\sum_{j=1}^{k} b_{ij} f_j$ and the specific components e_i do not undergo any change. The selection of a particular vector of factors, that is, the *identification* of the factors, requires additional criteria. For example, one of the factors has no impact on some of the variables or the sum of the squares of the loadings of one of the factors is maximum. Such constraints are discussed in Ref. 1 and they also depend on the particular application. There is a great variety of FA methods; we consider the following two to be the most important:

- *ML-Based FA*. If we have an $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{C})$ distributed sample, then we maximize its log-likelihood function

$$-\frac{1}{2} n \ln |\boldsymbol{C}| - \frac{1}{2} n \mathrm{tr}\ \boldsymbol{C}^{-1}\hat{\boldsymbol{C}} + \text{constant}$$

  with respect to $\boldsymbol{B}, \boldsymbol{D}$ subject to $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{D}$, where $|\boldsymbol{C}|$ is the determinant of $\boldsymbol{C}$ and $\hat{\boldsymbol{C}}$ the sample covariance matrix, estimated from an independent, identically distributed (iid) sample. To avoid the ambiguity due to rotation, we also put the constraint that $\boldsymbol{B}'\boldsymbol{D}^{-1}\boldsymbol{B}$ is diagonal. Equivalently, we have to solve

$$\ln |\boldsymbol{B}\boldsymbol{B}' + \boldsymbol{D}| + \mathrm{tr}\ (\boldsymbol{B}\boldsymbol{B}' + \boldsymbol{D})^{-1}\hat{\boldsymbol{C}} \to \min, \text{ subject to } \boldsymbol{B}'\boldsymbol{D}^{-1}\boldsymbol{B} \text{ diagonal}$$

  There are both theoretical results and algorithms based on numerical methods at our disposal to treat this problem, see Refs 1, 2.
- *PC-Based FA*. If the variance of e_i does not depend on i, that is, $\boldsymbol{D} = \sigma^2 \boldsymbol{I}_n$ with some $\sigma > 0$, then the columns of $\boldsymbol{B}$ span the same linear space as the first k eigenvectors of $\boldsymbol{C}$ do. This is the rationale for using the first k PCs of $\mathbf{X}$ to estimate the factors and the factor loadings, a widespread though unwarranted practice. Actually, Equation (1) yields $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{U}\mathbf{Y} = \boldsymbol{\mu} + (\boldsymbol{U}\boldsymbol{\Lambda}^{1/2})(\boldsymbol{\Lambda}^{-1/2}\mathbf{Y})$, that gives rise to estimate the factor loading matrix with $(\sqrt{\hat{\lambda}_1}\hat{\mathbf{u}}_1, \ldots, \sqrt{\hat{\lambda}_k}\hat{\mathbf{u}}_k)$, where $\hat{\lambda}_i$s and $\hat{\mathbf{u}}_i$s are the first k eigenvalues and eigenvectors of $\hat{\boldsymbol{C}}$ and k is selected according to the spectral gap of $\hat{\boldsymbol{C}}$.

As an example from meteorology, suppose that X_is are the yearly variations of average temperatures, observed in $n = 30$ European cities for $N = 60$ years (this is a sample). The factor structure above, with $k = 1$, would explain such a variation as depending on one common stochastic latent variable, plus local variables that have zero covariance with one another. However, FA was developed by psychometricians in the first half of the twentieth century (Spearman, Thurstone) and was used to find latent common factors behind rvs corresponding to results of psychological tests. The very meaning of the factors, like general intelligence, was established by the experts, based on the loadings of the individual factors in the variables X_is. The interpretation of the factors is the most straightforward if each variable is loaded highly on at

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

most one factor, and if all the factor loadings are either large (in absolute value) or near zero, with few intermediate values. Then the variables can be divided into disjoint sets, each of which is associated with one factor, and some variables may be left over. The factor f_j can be interpreted as the common feature of those X_is for which b_{ij} is large. We can make advantage of a $k \times k$ rotation $\boldsymbol{Q}$ such that the factor loading matrix $\boldsymbol{BQ}^{-1}$ is the best interpretable in the abovementioned sense. For this convenience, there are methods of rotation elaborated, for example, the VARIMAX rotation, see Ref. 2 for details.

2.3 Independent Component Analysis (ICA)

This is a more recent generalization of factoring methods for finding indeed independent and not only uncorrelated components behind the variables. PCA is mainly applicable to multivariate Gaussian rvs; in this case, the PCs are not only uncorrelated, but they are also independent. On the contrary, given the n-dimensional random vector $\mathbf{X}$ (assume that it is already "whitened," i.e., $\mathrm{Var}(\mathbf{X}) = \boldsymbol{I}_n$), ICA looks for an orthogonal transformation, obtained by an $n \times n$ orthogonal matrix $\boldsymbol{V}$ such that the *mutual information* of the components of the random vector $\mathbf{Y} = \boldsymbol{V}\mathbf{X}$ is minimized. It is defined as

$$I(\mathbf{Y}) = \sum_{i=1}^{n} H(Y_i) - H(\mathbf{Y}) = \sum_{i=1}^{p} H(Y_i) - H(\mathbf{X}) - \log_2 |\boldsymbol{V}| = \sum_{i=1}^{n} H(Y_i) - H(\mathbf{X}) \tag{6}$$

where $|\boldsymbol{V}| = 1, H(\mathbf{X}) = \mathbb{E}(-\log_2 f(\mathbf{X})) = -\int f(\mathbf{x}) \log_2 f(\mathbf{x})\, \mathrm{d}\mathbf{x}$ is the *differential entropy* of the continuously distributed rv $\mathbf{X}$ with probability density function f. Note that $I(\mathbf{Y})$ is also the Kullback–Leibler distance between the distribution of $\mathbf{Y}$ and its independent version, which is the product of its marginals. Therefore, $I(\mathbf{Y})$ is always nonnegative and is zero if and only if the Y_is are independent, see Ref. 3.

In view of Equation (6), as $H(\mathbf{X})$ is fixed, the sum of the entropies of Y_is should be minimized. Note that, among the absolutely continuous distributions with equal variance, the Gaussian distribution has the largest entropy. Therefore, we are looking for the orthogonal transformation $\boldsymbol{V}$ that maximizes the departure of the distribution of Y_is from the Gaussianity. There are algorithms to do so, using the *negentropy* measure for this departure or sample entropy estimation.

ICA makes sense if $\mathbf{Y}$ follows a non-Gaussian multivariate distribution, where independence is much stronger than pairwise uncorrelatedness. If $\mathbf{X}$ were multivariate Gaussian, then the usual PCA would result in a multivariate Gaussian $\mathbf{Y}$ with independent components and $I(\mathbf{Y}) = 0$ that cannot be further decreased. Also note that whitening $\mathbf{X}$ means a preliminary PCA-based FA on it, and via the selection of the orthogonal $\boldsymbol{V}$, ICA can as well be viewed as a method of factor rotation. For further aspects, see Ref. 4.

3 Dynamic Factor Analysis

3.1 Nonparametric View

Standard FA can be generalized to the case of an n-dimensional vector stochastic process $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots X_n(t))'$. Here, $t \geq 0$ is the time, and our sample usually consists of observations at discrete time instances $t = 1, \ldots, T$. In the classical FA approach, the data came from iid observations, and the dimension reduction happened in the so-called cross-sectional dimension, that is, the number n of variables was decreased. In dynamic FA, the observations $\mathbf{X}(t)$s are usually not iid, and we want to compress the information, embodied by them, in the cross-sectional and the time dimension as well. Sometimes, even the cross-sectional dimension n is large compared to the time dimension T.

DOI: 10.1002/9781118445112.stat07505.pub2

04554

Assume that $\mathbf{X}(t)$ is *weakly stationary* (*see* **Stationary Processes**) with an absolutely continuous spectrum, that is, $\mathbf{X}(t)$ has an $n \times n$ spectral density matrix $\mathbf{\Sigma}_{\mathbf{X}}$. With the integer $1 \leq k < n$, the *dynamic k-factor model* for $\mathbf{X}(t)$ is

$$\mathbf{X}(t) = \boldsymbol{\mu} + \boldsymbol{B}(L)\mathbf{f}(t) + \mathbf{e}(t) = \boldsymbol{\mu} + \chi(t) + \mathbf{e}(t) \tag{7}$$

or with components,

$$X_i(t) = \mu_i + b_{i1}(L)f_1(t) + \cdots + b_{ik}(L)f_k(t) + e_i(t)$$

where the k-dimensional stochastic process $\mathbf{f}(t) = (f_1(t), \ldots f_k(t))'$ is the *dynamic factor*, $\chi(t)$ is called *common component*, the n-dimensional stochastic process $\mathbf{e}(t) = (\mathbf{e}_1(t), \ldots, e_n(t))^{\mathrm{T}}$ is called *noise component*, and the $n \times k$ matrix $\boldsymbol{B}(L) = (b_{ij}(L))$, $i = 1, \ldots, n$, $j = 1, \ldots, k$, is the *transfer function*. Here, L is the *lag operator* and $b_{ij}(L)$ is a square-summable one-sided filter, that is, $b_{ij}(L) = b_{ij}(0) + b_{ij}(1)L + b_{ij}(2)L^2 + \ldots$ with $\sum_{\ell=0}^{\infty} b_{ij}^2(\ell) < \infty$. Further, the components of Equation (7) satisfy the following requirements:

$$\begin{aligned} \mathbb{E}(\mathbf{f}(t)) = \mathbf{0}, \quad & \mathbb{E}(\mathbf{e}(t)) = \mathbf{0}, \quad t \in \mathbb{Z} \\ \mathrm{Cov}(e_i(t), f_j(s)) = 0, \quad & i = 1, \ldots, n, \quad j = 1, \ldots, k, \quad t, s \in \mathbb{Z},\ s \leq t \\ \mathrm{Cov}(e_i(t), e_j(s)) = 0, \quad & i, j = 1, \ldots n, \quad i \neq j, \quad t, s \in \mathbb{Z},\ s < t \end{aligned} \tag{8}$$

$\mathbf{f}(t)$ and $\mathbf{e}(t)$ are also weakly stationary and they have spectral densities $\mathbf{\Sigma}_{\mathbf{f}}$ and $\mathbf{\Sigma}_{\mathbf{e}}$, the model Equation (7) extends to the spectral density matrices:

$$\mathbf{\Sigma}_{\mathbf{X}}(\theta) = \boldsymbol{B}(\mathrm{e}^{-i\theta})\mathbf{\Sigma}_{\mathbf{f}}(\theta)\boldsymbol{B}(\mathrm{e}^{-i\theta})^* + \mathbf{\Sigma}_{\mathbf{e}}(\theta), \quad \theta \in [-\pi, \pi] \tag{9}$$

where i is the imaginary unit and * denotes the complex conjugation. Here, we use the representation theory of weakly stationary processes (as the first and second moments uniquely define the process, we can use Hilbert spaces for the representation). In this setup, the time-lag can be identified with a unitary matrix, and the spectral theory of unitary matrices can be used (*see* **Spectral Methods**).

07422
04554

Very frequently, $\mathbf{f}(t)$ is assumed to be an orthonormal *white noise*(*see* **Stationary Processes**). Then Equation (9) simplifies to

$$\mathbf{\Sigma}_{\mathbf{X}}(\theta) = \boldsymbol{B}(\mathrm{e}^{-i\theta})\boldsymbol{B}(\mathrm{e}^{-i\theta})^* + \mathbf{\Sigma}_{\mathbf{e}}(\theta) \tag{10}$$

The so-called static case occurs if, in addition, $\boldsymbol{B}$ is constant (in time). Then with the covariance matrices, which by slightly abusing the notation are also denoted by $\mathbf{\Sigma}$ (there is indeed a one-to-one correspondence between them and the spectral density matrices), Equation (10) boils down to

$$\mathbf{\Sigma}_{\mathbf{X}} = \boldsymbol{B}\boldsymbol{B}' + \mathbf{\Sigma}_{\mathbf{e}}$$

that resembles Equation (5). Otherwise, Equation (7) is dynamic in that the latent variables $f_j(t)$s can affect the observables $X_i(t)$s both contemporaneously and with lags. Using the same example as in the previous section, here different cities may load the factors with different dynamics, that is, with different loading polynomials. In case of macroeconomic data, the latent dynamic factors are the driving forces of the underlying economic stochastic process, describing comovements between the multidimensional and strongly coherent time series, represented by the variables. It is also noted that estimation of the dynamic model requires estimation of both the contemporaneous and lagged covariances. Thus, the dynamic model, which yields a richer representation as compared to the standard one, can only be applied when a sufficiently large data set is available. Like in the standard factor model, neither $\boldsymbol{B}(L)$ nor $\mathbf{f}(t)$ are identified uniquely;

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

and given the spectral density $\mathbf{\Sigma}_{\mathbf{X}}$, the spectra $\mathbf{\Sigma}_{\boldsymbol{\chi}}$ and $\mathbf{\Sigma}_{\mathbf{e}}$ are generically unique for $k \leq n - \sqrt{n}$ (reminiscent of the Lederman bound). In addition, similar to the standard model, neither the common nor the specific component change if the factors and the transfer functions are exposed to a regular linear (dynamic) transformation.

We remark that if conditions are not imposed on the filters $b_{ij}(L)$s or on the processes $\mathbf{e}_i(t)$, the model is nonparametric. Interesting parametric specifications are obtained by assuming that the vector $\mathbf{e}(t)$ is white noise and that the filters $b_{ij}(L)$ are finite. See Section 3.3 for this approach.

In the nonparametric setting, the spectral densities can be estimated by ML, using Equation (9) frequency by frequency, that is, using the method employed to estimate Equation (5). Very often, in addition to the general assumptions (Equation 8), it is assumed that the noise components are uncorrelated, that is, that $\mathbf{\Sigma}_{\mathbf{e}}$ is diagonal, in which case, $\mathbf{e}(t)$ is called *idiosyncratic noise*. In Ref. 5, it was proved that for dynamic factor models with idiosyncratic noise, the set of all spectral densities $\mathbf{\Sigma}_{\mathbf{X}}(\theta)$ described by Equation (9) for given k, is a "thin" subset of the set of all spectral densities $\mathbf{\Sigma}_{\mathbf{X}}(\theta)$ if $k < n - \sqrt{n}$ holds. Our purpose is to separate this noise component from the common component $\boldsymbol{\chi}(t)$ of Equation (7). More specifically, if $\mathbf{\Sigma}_{\mathbf{e}}(\theta) = \sigma_{\mathbf{e}}^2(\theta)\boldsymbol{I}_n$, the spectral density $\boldsymbol{B}(\mathrm{e}^{-i\theta})\mathbf{\Sigma}_{\mathbf{f}}(\theta)\boldsymbol{B}(\mathrm{e}^{-i\theta})^*$ can be estimated by means of Brillinger's **principal components** in the frequency domain, see Ref. 6. This is, in fact, a PCA technique, where we approach the n-dimensional observed process $\mathbf{X}(t)$ by a filtered version of itself, whose spectral density is of reduced rank ($k < n$), such that the variance of the residuals is minimized. Akin to the PCA model, here the number of factors k is given and not a property of $\mathbf{\Sigma}_{\mathbf{X}}$[7−10].

In the following sections, we discuss estimation issues in both a nonparametric high-dimensional and a parametric framework.

3.2 Generalized Dynamic Factor Model (GDFM)

It is often the case that the number of **Time Series** available is too large to estimate a dynamic factor model by ML. It can be proved that when the data set is large, PCA is a valid alternative to ML. More precisely, under suitable assumptions, as both T (the number of observations) for each time series and n (the number of time series) tend to infinity, PCs of the observable variables $X_i(t)$s provide consistent estimators of the common and specific components.

In a nutshell, the basic idea behind high-dimensional dynamic factor models is that when $n \to \infty$, we can get rid of the specific components by averaging. As we will see below, the restriction that the specific components are uncorrelated can even be relaxed so that a "weak" correlation between them is allowed. Moreover, arithmetic averages are fruitfully replaced by PCs.

Indeed, the assumption that the specific components are uncorrelated is also too restrictive for many applications., where "local" dependency between the noise components may occur. Moreover, in a number of applications, for example, in cross-country business cycle analysis, asset pricing, or monitoring and forecasting economic activity by estimation of common factors (diffusion indices), the cross-sectional dimension n may be high, possibly exceeding the sample size, see, for example, Ref. 11.

In the subsequent general model, the cross-sectional dimension n is not fixed, and we consider the double sequence $\{X_i(t) : i \in \mathbb{N},\ t \in \mathbb{Z}\}$, where the assumption of Equation (7) holds for every vector $\mathbf{X}^n(t) = (X_1(t), \ldots, X_n(t))'$ with $n \in \mathbb{N}$. Thus, we have a sequence of GDFM's (generalized dynamic factor models)

$$\mathbf{X}^n(t) = \boldsymbol{B}^n(L)\mathbf{f}(t) + \mathbf{e}^n(t) = \boldsymbol{\chi}^n(t) + \mathbf{e}^n(t), \quad n \in \mathbb{N} \tag{11}$$

where the expectations are zeros and the noise vectors $\mathbf{e}^n(t)$s and the transfer functions $\boldsymbol{B}^n(L)$s are nested; further, the k-dimensional dynamic factor $\mathbf{f}(t)$ and the dimension k itself do not depend on n. We assume

DOI: 10.1002/9781118445112.stat07505.pub2

that the stochastic process $\mathbf{X}^n(t)$ is weakly stationary and has a spectral density (matrix) $\mathbf{\Sigma}_{\mathbf{X}}^n(\theta)$, for $n = 1, 2, \ldots$ in a nested way. As $\mathbf{\Sigma}_{\mathbf{X}}^n(\theta)$ is Hermitian, its eigenvalues are nonnegative reals, let us denote them by $\lambda_{\mathbf{X},1}^n(\theta) \geq \cdots \geq \lambda_{\mathbf{X},n}^n(\theta)$, which are, in fact, $[-\pi, \pi] \to \mathbb{R}_+$ functions. Accordingly, $\mathbf{\Sigma}_{\mathbf{e}}^n(\theta)$ will denote the spectral density matrix of $\mathbf{e}^n(t)$ and $\mathbf{\Sigma}_{\boldsymbol{\chi}}^n(\theta)$ that of $\boldsymbol{\chi}^n(t)$.

Weak dependence between the noise components of $\mathbf{e}^n(t)$ (instead of the classical assumption of uncorrelatedness) is formalized by requiring the largest eigenvalue of the spectral density $\mathbf{\Sigma}_{\mathbf{e}}^n(\theta)$ to be uniformly bounded for all n. At the same time, the first k eigenvalues of $\mathbf{\Sigma}_{\boldsymbol{\chi}}^n(\theta)$ diverge almost everywhere in $[-\pi, \pi]$ as $n \to \infty$. Of course, the conditions of the forthcoming representation result apply to the spectrum of the observed sequence of time series $\mathbf{X}^n(t)$.

Before stating this result, we again emphasize that the basic idea in dealing with GDFMs 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 GDFM with $\boldsymbol{B}^n = (1, 1, \ldots, 1)^{\mathrm{T}}$, $f(t)$ iid and $e_i(t)$ iid in i and t. Then, as $\frac{1}{n}\sum_{i=1}^n e_i(t)$ converges to zero as $n \to \infty$, the noise term is averaged out.

Forni and Lippi[12] provide necessary and sufficient conditions for the existence of an underlying GDFM in terms of the observable spectral densities $\mathbf{\Sigma}_{\mathbf{X}}^n(\theta)$, $n \in \mathbb{N}$. Here, we formulate the theorem as is stated in Ref. 13.

Theorem 2. *The double sequence* $\{\mathbf{X}^n(t) : n \in \mathbb{N}, t \in \mathbb{Z}\}$ *can be represented by a sequence (Equation 11) of GDFMs if and only if*

- *the first k eigenvalues,* $\lambda_{\mathbf{X},1}^n(\theta) \geq \cdots \geq \lambda_{\mathbf{X},k}^n(\theta)$ *(in nonincreasing order), of* $\mathbf{\Sigma}_{\mathbf{X}}^n(\theta)$ *diverge almost everywhere in* $[-\pi, \pi]$ *as* $n \to \infty$*;*
- *the* $(k+1)$*th eigenvalue* $\lambda_{\mathbf{X},k+1}^n(\theta)$ *of* $\mathbf{\Sigma}_{\mathbf{X}}^n(\theta)$ *is uniformly bounded for* $\theta \in [-\pi, \pi]$ *almost everywhere and for all* $n \in \mathbb{N}$*.*

In Ref. 14, the authors show that the sequence of PCA models for $\mathbf{X}^n(t)$, $n \in \mathbb{N}$, approximates the corresponding sequence of GDFMs in probability as $n, T \to \infty$, see also Ref. 13 for details. We saw that in case of the classical PCA and FA, the PCs do not provide a consistent estimator of the factor structure unless the specific components have the same variance. However, in the high-dimensional dynamic factor model, even when the variances of the specific components are not equal and even when the covariance between them is not zero, the covariances of $X_i(t)$s are increasingly determined by the covariance of the common components as $n \to \infty$.

Note that in empirical situations, the number of factors is not known. For further information on this issue, see Refs 15 and 16. The high-dimensional standard factor model was first introduced in Refs 17 and 18; for extensions to the dynamic case, see Refs 12 and 19.

3.3 Parametric View

Here, under a given number of factors and a finite lag parameter, factor loadings are estimated, while an autoregression is fitted for the factor process. We present an improved version of the model[20], where the authors introduced an iteration that uses regression and PCA methods to find the factors one by one; the method is based on the work of Box and Tiao[7] using canonical transformations for multiple time series. The improved version[21] is able to extract dynamic factors simultaneously, rather than sequentially.

The input data are n-dimensional observations $\mathbf{X}(t) = (X_1(t), \ldots, X_n(t))'$, where t is the time and the process is observed at equidistant instances between T_1 and T_2. For a given positive integer $k < n$, we are looking for (at all leads) uncorrelated factors $f_1(t), \ldots, f_k(t)$ such that they satisfy the following model equations.

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

1. The first one is the linear model:

$$f_j(t) = \sum_{i=1}^{n} b_{ij} X_i(t), \qquad t = T_1, \ldots, T_2; \quad j = 1, \ldots, k \tag{12}$$

2. The second one is the dynamic equation of the factors:

$$\hat{f}_j(t) = c_{j0} + \sum_{\ell=1}^{L} c_{j\ell} f_j(t-\ell), \qquad t = T_1 + L, \ldots, T_2; \quad j = 1, \ldots, k \tag{13}$$

   where the lag length L is a given positive integer and $\hat{f}_j(t)$ is the Lth order auto-regressive prediction of the jth factor at date t.
3. The third one is the linear prediction of the variables by the factors:

$$\hat{X}_i(t) = d_{0i} + \sum_{j=1}^{k} d_{ji} f_j(t), \qquad t = T_1, \ldots, T_2; \quad i = 1, \ldots, n \tag{14}$$

The parameters of the model in matrix form are $\boldsymbol{B} = (b_{ij})$, $\boldsymbol{C} = (c_{j\ell})$, $\boldsymbol{D} = (d_{ji})$, $i = 1, \ldots, N, j = 1, \ldots, k$, $\ell = 1, \ldots L$, where the estimates of the parameters c_{j0}, d_{0i} can be expressed in terms of these. The parameters are estimated so that the objective function

$$w_0 \sum_{j=1}^{k} \mathrm{Var}_L(f_j - \hat{f}_j) + \sum_{i=1}^{n} w_i \mathrm{Var}(X_i - \hat{X}_i) \tag{15}$$

is minimized on the conditions for the orthogonality and variance of the factors:

$$\mathrm{Cov}(f_i, f_j) = 0, \quad i \neq j; \qquad \mathrm{Var}(f_j) = v_j, \quad j = 1, \ldots, k \tag{16}$$

In Equation (15), the subscript L indicates that the time variation is restricted to dates $T_1 + L, \ldots, T_2$ only, $w_0, w_1, \ldots, w_n$ are given non-negative constants (balancing between the dynamic and static part), while the positive numbers v_js are the variances of the individual factors indicating their relative importance.

To estimate the parameters, the covariances between the original and the lagged time series are used. The estimated covariances between the X_is are collected in the $n \times n$ empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{X}} = (\sigma_{ij})$. The lagged time series $\mathbf{Z}^j(t)$ for $j = 1, \ldots, L$ are defined with the following coordinates:

$$Z_i^j(t) = X_i(t) - \sum_{\ell=1}^{L} c_{j\ell} X_i(t-\ell), \quad t = T_1 + L, \ldots, T_2; \quad i = 1, \ldots, n \tag{17}$$

Let $\hat{\boldsymbol{\Sigma}}_{\mathbf{Z}^j}$ denote the $n \times n$ symmetric, positive semidefinite empirical covariance matrix of the lagged time series $\mathbf{Z}^j(t)$ of the following entries:

$$\mathrm{Cov}_L(Z_i^j, Z_m^j) = \frac{1}{T_2 - T_1 - L + 1} \sum_{t=t_1+L}^{T_2} (Z_i^j(t) - \overline{Z}_i^j)(Z_m^j(t) - \overline{Z}_m^j) \tag{18}$$

where $\overline{Z}_i^j = \frac{1}{T_2 - T_1 - L + 1} \sum_{t=T_1+L}^{T_2} Z_i^j(t)$, $i = 1, \ldots, n, j = 1, \ldots, k$.

DOI: 10.1002/9781118445112.stat07505.pub2

With these, the objective function (15) to be minimized is

$$G(\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}) = w_0 \sum_{j=1}^{k} \mathbf{b}_j^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\mathbf{Z}^j} \mathbf{b}_j + \sum_{i=1}^{n} w_i \sigma_{ii} - 2 \sum_{i=1}^{n} w_i \sum_{j=1}^{k} d_{ji} \sum_{m=1}^{n} b_{mj} \sigma_{im} + \sum_{i=1}^{n} w_i \sum_{j=1}^{k} d_{ji}^2 v_j$$

where $\mathbf{b}_j \in \mathbb{R}^n$ is the jth column of the matrix $\boldsymbol{B}$, and the minimum is taken on the constraints

$$\mathbf{b}_j^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\mathbf{X}} \mathbf{b}_h = \delta_{jh} v_j, \qquad j,h = 1, \ldots, k \tag{19}$$

The procedure finding the minimum is based on the following iteration that consists of an outer and an inner cycle. Choosing an initial $\boldsymbol{B}^{(0)}$ of columns satisfying Equation (19), the following three steps are alternated in the itth outer iteration.

1. Starting with $\boldsymbol{B}^{(it)}$ we calculate the f_js based on Equation (12), then we fit a li model to estimate the parameters of t toregression (Equation 13). Hence, the current value is obtained.
2. On the basis of this , we find matrices $\hat{\boldsymbol{\Sigma}}_{\mathbf{Z}^j}$ using Equations (17) and (18), $j = 1, \ldots, k$. Putting this auxiliary variable into $G(\boldsymbol{B}^{(it)}, \boldsymbol{C}^{(it)}, \boldsymbol{D})$, we take its minimum with respect to $\boldsymbol{D}$, while keeping $\boldsymbol{B}$ and $\boldsymbol{C}$ fixed. This is done by differentiation. Say, the minimum is taken at $\boldsymbol{D}^{(it)}$.
3. Now keeping $\boldsymbol{C}$ and $\boldsymbol{D}$ fixed, we minimize $G(\boldsymbol{B}, \boldsymbol{C}^{(it)}, \boldsymbol{D}^{(it)})$ with respect to $\boldsymbol{B}$. This minimization needs an inner cycle. The $\boldsymbol{B}$, at which the minimum is attained, will be the next $\boldsymbol{B}^{(it+1)}$.

With this new $^{1)}$, we return to Step 1 of the outer cycle ($it := it + 1$) and proceed until convergence. As the value of the nonnegative objective function is in each step decreased, the convergence of the algorithm to a local minimum can be guaranteed.

In the inner cycle, we can write $G(\boldsymbol{B}, \boldsymbol{C}^{(it)}, \boldsymbol{D}^{(it)})$ as the sum of k inhomogeneous quadratic forms, which is maximized under orthogonality constraints for the transformed $\mathbf{b}_j$s. This problem is solved in Ref. 22 via finding *compromise systems* of symmetric matrices.

As an application, the authors[21] extracted $k = 3$ factors out of $n = 10$ yearly aggregated Hungarian macroeconomic indicators spanning 1993–2007 with $L = 4$. As the variables were measured in different units, they were normalized; and the reciprocals of their standard deviations were used as weights $w_1, \ldots, w_n$ in the objective function (15). In view of Ref. 20, the authors used the same weights $v_j = T_2 - T_1 + 1$ ($j = 1, \ldots, k$) for the factors and the choice $w_0 = n/kv_1$, ensuring the equilibrium between the dynamic and static parts.

Figures 1–3 illustrate the dynamic factors in time. The first two factors show some tendency starting from the year 2000, while the third factor is somewhat antipodal to the first one. Possibly, only the first two factors are significant, while the next ones are dampened dummies of them. We remark that in our model $k = 3$ is, in fact, the maximum number of factors, which does not contradict to certain rank conditions. The actual number of factors can be less, depending on the least square errors and practical considerations; it is an expert's job to decide how many factors to retain.

The coefficients of the matrices $\boldsymbol{B}$, $\boldsymbol{D}$, and $\boldsymbol{C}$ are shown in Tables 1–3, respectively. The relatively high constant terms in the linear prediction of the components by the factors (Table 2) refer to "small" communalities. However, the constant coefficients in the autoregressive model are small (Table 3) and the coefficient belonging to lag 2 is the largest in all the three factors. Note that since 1990, there are national

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

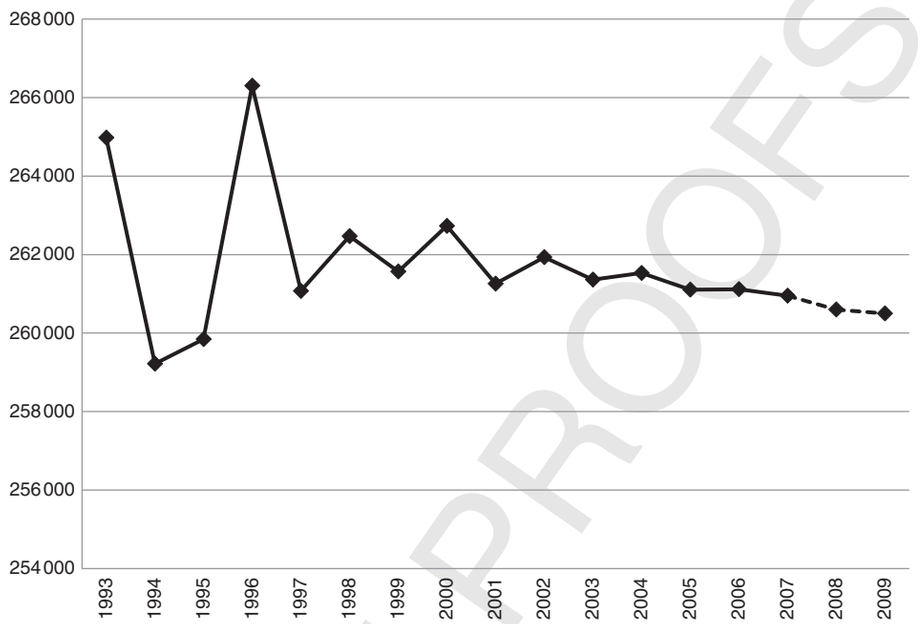

Figure 1. Dynamic factor 1.

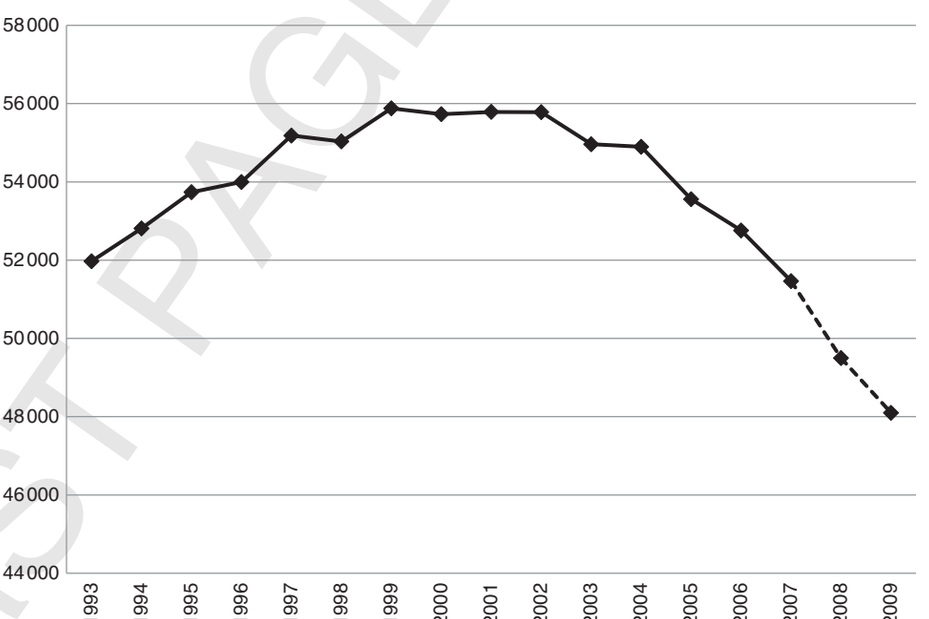

Figure 2. Dynamic factor 2.

DOI: 10.1002/9781118445112.stat07505.pub2

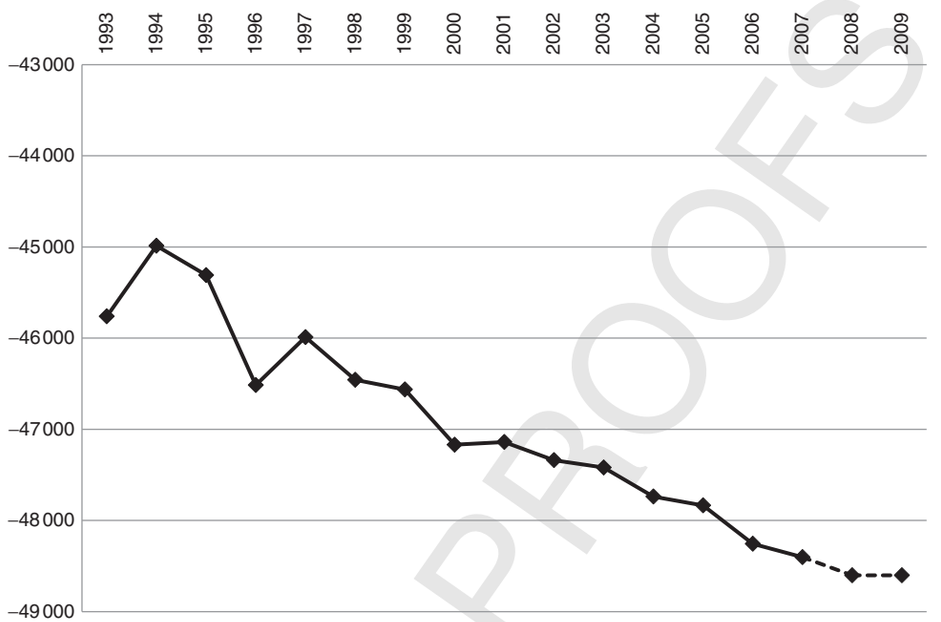

Figure 3. Dynamic factor 3.

Table 1. Factors expressed in terms of the components (matrix ***B***).

	Factor 1	Factor 2	Factor 3
GDP	38.324	−2.541	−6.116
EDU	−1.775	5.725	0.015
HEALTH	10.166	0.837	−1.650
IND	−0.261	0.255	−0.107
AGR	6.146	2.919	−1.124
ENERGY	24.082	4.592	−4.054
IMP	1.560	−1.209	−0.213
EXP	−3.907	−0.233	0.615
INV	2.864	0.038	−0.510
INNOV	−0.608	0.197	0.089

elections in every fourth year in Hungary, and lag 2 may correspond to the mid-period, when the measures introduced by the new government could have the higher impact on the economy.

Predictions were also made for 2 years ahead by means of the matrix ***C***. The predicted factor values for 2008 and 2009 are illustrated by dashed lines and they show decline in all the three factors, possibly indicating the upcoming economic crisis. The 2008s estimates prohibited a good fit to the factual data in case of most variables with squared error 1.16 of this only year that is comparable to the MSE 0.82 based on 15 years.

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

Table 2. Components estimated by the factors (matrix ***D***).

	Factor 1	Factor 2	Factor 3	Constant
GDP	−0.108	−0.025	−0.677	−0.670
EDU	−0.142	0.145	−0.877	−8.637
HEALTH	0.115	−0.132	0.656	16.250
IND	−0.898	−0.187	−5.784	−14.690
AGR	0.021	0.005	0.137	6.809
ENERGY	0.085	−0.038	0.543	10.055
IMP	−0.098	−0.152	−0.868	0.311
EXP	−0.516	−0.931	−1.840	109.915
INV	−0.209	0.026	−1.341	−6.779
INNOV	−0.061	0.121	−0.484	−9.867

Table 3. Dynamic equations of the factors (matrix ***C***).

Lag	Factor 1	Factor 2	Factor 3
0	−0.000	0.001	−0.000
1	0.069	0.283	0.117
2	0.473	1.644	0.495
3	0.205	0.229	0.141
4	0.251	−1.168	0.258

4 Conclusion

The basic ideas underlying small- and medium-sized standard and dynamic factor models have been reviewed. Maximum likelihood is the main estimation tool, whereas the method of PCs is confined to the special case of equal-variance- specific components.

GDFMs allow for a more general structure for the covariance of the specific components, with zero covariance between them being no longer necessary. Moreover, as both T and n diverge, PCs, either in the time or in the frequency domain, provide consistent estimators for the common and specific components.

High-dimensional dynamic factor models exhibit a very general structure, only requiring stationarity either directly or after a suitable transformation of the data. They are, therefore, a promising tool of analysis for many empirical situations in which a large evolving data set is available.

A parametric model is also presented which is capable of parameter estimation in small-sized macroeconomic problems via techniques of linear models and autoregression. An algorithm is also presented that uses compromise systems of inhomogeneous quadratic forms and is illustrated via a numerical example.

01663
02354
02443
02457
03025
03440
04554
07372
07422
07702
07813

Related Articles

Maximum Likelihood Estimation; **Spectral Decomposition**; **Factor Analysis**; **Principal Components**; **Stochastic Processes**; **Econometrics**; **Stationary Processes**; **Covariance Matrix Estimation (Estimated Parameters)**; **Spectral Methods**; **Forecasting, Environmental**; **Space-Time Kalman Filter**.

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

References

[1] Lawley, D.N. and Maxwell, A.E. (1971) *Factor Analysis as a Statistical Method*, 2nd edn, Butterworths, London.
[2] Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979) *Multivariate Analysis*, Academic Press, London.
[3] Cover, T.M. and Thomas, J.A. (2005) *Elements of Information Theory*, 2nd edn, John Wiley & Sons, Inc., New York.
[4] Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Springer-Verlag, New York.
[5] Scerrer, W. and Deistler, M. (1998) A structure theory for linear dynamic errors-in-variables models. *SIAM J. Control Optim.* **36** (6), 2148–2175.
[6] Brillinger, D.R. (1981) *Time Series Data Analysis and Theory*, Holden Day, San Francisco, CA.
[7] Box, G.E.P. and Tiao, G.C. (1977) A canonical analysis of multiple time series. *Biometrika* **64** (2), 355–365.
[8] Peña, D. and Box, G.E.P. (1987) Identifying a simplifying structure in time series. *J. Am. Stat. Assoc.* **82**, 836–843.
[9] Sargent, T.J. and Sims, C.A. (1977) Business cycle modelling without pretending to have too much a priori economic theory, in *New Methods in Business Cycle Research* (ed. C.A. Sims), Federal Reserve Bank of Minneapolis, Minneapolis, MN, pp. 45–109.
[10] Geweke, J. (1977) The dynamic factor analysis of economic time series, in *Latent Variables in Socio-Economic Models* (eds D.J. Aigner and A.S. Goldberger), North Holland, Amsterdam, pp. 365–383.
[11] Stock, J.H. and Watson, M.W. (2002) Macroeconomic forecasting using diffusion indexes. *J. Bus. Econ. Stat.* **20**, 147–162.
[12] Forni, M. and Lippi, M. (2001) The generalized dynamic factor model: representation theory. *Economet. Theory* **17**, 1113–1341.
[13] Deistler, M. and Zinner, C. (2007) Modelling high-dimensional time series by generalized linear dynamic factor models: an introductory survey. *Commun. Inf. Syst.* **7** (2), 153–166.
[14] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000) The generalized dynamic factor model: identification and estimation. *Rev. Econ. Stat.* **82**, 540–554.
[15] Bai, J. and Ng, S. (2002) Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.
[16] Hallin, M. and Livska, R. (2007) Determining the number of factors in the general dynamic factor model. *J. Am. Stat. Assoc.* **102**, 603–617.
[17] Chamberlain, G. (1983) Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51**, 1281–1304.
[18] Chamberlain, G. and Rothschild, M. (1983) Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305–1324.
[19] Stock, J.H. and Watson, M.W. (2002) Forecasting using principal components from a large number of predictors. *J. Am. Stat. Assoc.* **97**, 1167–1179.
[20] Bánkövi, G., Veliczky, J., and Ziermann, M. (1983) Estimating and forecasting dynamic economic relations on the basis of multiple time series. *Z. Angew. Math. Mech.* **63**, 398–399.
[21] Bolla, M. and Kurdyukova, A. (2010) Dynamic factors of macroeconomic data. *Ann. Univ. Craiova* **37**, 1–11.
[22] Bolla, M., Michaletzky, Gy., Tusnády, G., and Ziermann, M. (1998) Extrema of sums of heterogeneous quadratic forms. *Linear Algebra Appl.* **269**, 331–365.

Further Reading

Anderson, B. and Deistler, M. (2008) Generalized Linear Dynamic Factor Models – A Structure Theory. 2008 IEEE Conference on Decision and Control.
Anderson, B. and Deistler, M. (2008) Properties of zero-free transfer function matrices. *SICE J. Control Meas. Syst. Integr.* **1**, 1–9.
Bernanke, B.S. and Boivin, J. (2003) Monetary policy in a data-rich environment. *J. Monetary Econ.* **50**, 525–546.
Bernanke, B.S., Boivin, J., and Eliasz, P. (2005) Measuring monetary policy: a factor augmented autoregressive (FAVAR) approach. *Q. J. Econ.* **120**, 387–422.
Denton, F.T. (1978) Single-equation estimators and aggregation restrictions when equations have the same set of regressors. *J. Economet.* **8**, 173–179.
Forni, M. and Gambetti, L. (2010) The dynamic effects of monetary policy: a structural factor model approach. *J. Monetary Econ.* **57**, 203–216.

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07505.pub2

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009) Opening the black box: structural factor models with large cross-sections. *Economet. Theory* **25**, 1319–1347.

Forni, M. and Lippi, M. (2011) The generalized dynamic factor model: one-sided representation results. *J. Economet.* **163**, 23–28.

Geweke, J.F. and Singleton, K.J. (1981) Maximum likelihood "confirmatory" factor analysis of economic time series. *Int. Econ. Rev.* **22**, 37–54.

DOI: 10.1002/9781118445112.stat07505.pub2