Introduction to Non-Stationary Dynamic Factor Models Estimated by Spectral PCA with Conjugate 2D-DFT

i4cast LLC

Initial Version: November 2025

This Version: November 2025

Abstract

This paper introduces an estimation of non-stationary dynamic factor models (non-stationary DFMs) by spectral principal components analysis (spectral PCA) in frequency domain with "conjugate two-dimensional" discrete Fourier transform (conjugate 2D-DFT). Additional PCA in time domain and classic DFT are also used. This method can estimate large DFMs on big datasets of vector time-series. Using publicly available PCA and DFT codes, implementations of this DFM estimation framework can be scaled into robust machine learning algorithms to be used by public.

Keywords: Dynamic factor models, DFM, Non-stationary DFM, multivariate time-series analysis and forecast, spectral density matrix, spectral principal components analysis, spectral PCA, dynamic PCA, vector autoregression, vector autocovariance.

List of Sections

1. Introduction
2. Estimating DFM by Spectral PCA

3. DFM Equations

4. Vector Time-Series Forecasts

5. Shock Response Forecasts

6. Multivariate Volatility Forecasts

7. Performance Metrics and Benchmarks

8. Examples

9. Discussion

10. Further Development

A.1. Conjugate 2D-DFT

A.2. Eigenpairs of VCMs and SDMs

A.3. Spectral PCA

A.4. PCA after Inv-C2D-DFT

Reference

1. Introduction

Many explanatory or predictive analyses on big datasets can be carried out by modeling sets of large number of time-series. Most often, a large set of observed time-series are influenced by smaller number of unobserved common factors, with each factor affecting all or many time-series in the dataset. Factor models (FMs) are the most popular method applying to time-series datasets with underlying unobserved common factors.

Factor models and model estimations can be classified by various features as

- stationary: values of model coefficients are not changing over time,
- non-stationary (time-varying): values of model coefficients are changing over

time,

- static: model coefficients make values of time-series related only cross-sectionally but not over time,
- dynamic: model coefficients make values of time-series related both cross-sectionally and over time,
- frequentist: model coefficients are assumed as deterministic and unobserved components are treated as stochastic,
- Bayesian: all model coefficients and unobserved components are estimated as stochastic by Bayesian formulas,
- time domain: all variables at all model estimation steps are in time domain,
- frequency domain: variables in core estimation steps are in frequency domain, using Fourier transform to convert data between two domains.

Therefore, there are four categories of factor models:

1. stationary static factor models (stationary FMs),
2. non-stationary static factor models (non-stationary FMs),
3. stationary dynamic factor models (stationary DFMs),
4. non-stationary dynamic factor models (non-stationary DFMs).

And there are various types of factor model estimations:

- by frequentist approaches,
- by Bayesian analysis,
- by time-domain method,
- by frequency domain (spectral) analysis,

as well as other cases such as “by spectral analysis with Bayesian estimates”.

Specifically, a stationary static factor model can be expressed as

$$y_t = X_0 f_t + u_t , \tag{1.1}$$

where y_t is $n \times 1$ data vector representing n $(n \gg 1)$ observed time-series (with mean values removed), f_t is $m \times 1$ estimated vector representing m $(m < n$ or $m \ll n)$ unobserved common factor scores, X_0 is $n \times m$ estimated matrix representing $n \times m$ unobserved static factor loadings, and u_t is $n \times 1$ balancing vector representing n unobserved idiosyncratic random (error or noise) values. A non-stationary (time-varying) static factor model can be expressed as

$$y_t = X_{t,0} f_t + u_t , \tag{1.2}$$

where elements of $n \times m$ static factor loadings matrix $X_{t,0}$ are functions of time t.

A stationary dynamic factor model of order p $(p \geq 1)$ can be expressed as

$$y_t = X_0 f_t + X_1 f_{t-1} + \cdots + X_p f_{t-p} + u_t , \tag{1.3}$$

where X_j, $j = 0, 1, \cdots, p$, are $p + 1$ dynamic factor loadings matrixes, each composed of $n \times m$ constant loadings. Finally, a non-stationary dynamic factor model can be expressed as

$$y_t = X_{t,0} f_t + X_{t,1} f_{t-1} + \cdots + X_{t,p} f_{t-p} + u_t , \tag{1.4}$$

where elements of $p + 1$ dynamic factor loadings matrixes $X_{t,j}$ are functions of time t.

When estimated by frequentist approaches, unobserved factor loadings X_j or $X_{t,j}$ are assumed to be deterministic, while observed data y_t, unobserved factor scores f_t and unobserved errors u_t are treated as stochastic variables. When estimated with Bayesian formula, all y_t, X_j, $X_{t,j}$, f_t and u_t are estimated as stochastic variables.

Model coefficients, e.g. factor loadings and factor scores, of static and dynamic factor models in Eqs. (1.1) through (1.4) can be estimated by a variety of different techniques, as reviewed by Doz and Fuleky (2020). One of the methods discussed by Barigozzi (2024) and Doz and Fuleky (2020) is to estimate a stationary DFM by "dynamic (or spectral) principal components analysis" (DPCA or SPCA) in frequency domain through Fourier transform.

This paper is about estimating non-stationary dynamic factor model, Eq. (1.4), by spectral principal components analysis with frequentist approach.

Our time-varying DFM is developed to make (a) dynamic analysis on observed multiple time-series or "vector time-series", (b) multi-step forecasts of (multivariate) vector time-series, and (c) multi-step forecasts of multivariate volatility (covariance and vector autocovariance) of vector time-series.

2. Estimating DFM by Spectral PCA

Assume that we have a set of large number, n $(n \gg 1)$, of observed time-series (with mean values removed). We denote them collectively as time-series of $n \times 1$ observed data vector y_t. Each element of y_t represents an individual time-series. And assume that the vector time-series y_t follow a non-stationary vector autoregressive process as

$$y_t = M_{t,1}\, y_{t-1} + M_{t,2}\, y_{t-2} + \cdots + M_{t,p}\, y_{t-p} + e_t\ , \tag{2.1}$$

where $M_{t,j}$ are unobserved non-stationary $n \times n$ matrixes of vector autoregressive (VAR) coefficients, and e_t is unobserved $n \times 1$ stochastic vector composed of random disturbances, measurement errors and exogenous shocks. Total count of elements of all VAR matrixes in Eq. (2.1) is $n \times n \times p$ and could be a huge number.

When the data generating process of the observed vector time-series is assumed as having a smaller number, m $(m \ll n)$, of unobserved common factors, with each factor influencing all or many individual time-series dynamically (i.e. over time), the dynamic process of vector time-series y_t can be expressed with a more parsimonious dynamic factor model (DFM) in a state space representation as

$$y_t = X_{t,0}\, f_t + X_{t,1}\, f_{t-1} + \cdots + X_{t,\,p_1}\, f_{t-p_1} + u_t \ , \tag{2.2}$$

$$f_t = A_{t,1}\, f_{t-1} + A_{t,2}\, f_{t-2} + \cdots + A_{t,\,p_2}\, f_{t-p_2} + v_t \ , \tag{2.3}$$

where Eqs. (2.2) and (2.3) are measurement equation of order p_1 and state equation of order p_2, respectively. In the two state space equations above,

- time-series of $m \times 1$ unobserved vector f_t are common factors scores,
- non-stationary $n \times m$ matrixes $X_{t,j}$ are dynamic factor loadings measuring sensitivities of y_t on f_t,
- non-stationary $m \times m$ matrixes $A_{t,k}$ are vector autoregressive (VAR) coefficients of factor scores,
- time-series of $n \times 1$ vector u_t are idiosyncratic disturbances and measurement errors, and
- time-series of $m \times 1$ vector v_t are stochastic factor innovations and factor prediction errors.

There are many approaches estimating the factor loadings $X_{t,j}$, factor scores f_t and factor VAR coefficients $A_{t,k}$. One popular and practical way is to estimate these coefficients and values by principal components analysis (PCA) on covariance matrixes of observed data vector y_t,

$$C_{t,j,k} = Cov(\, y_{t-j} \,,\, y_{t-k} \,) = E(\, y_{t-j}\, y_{t-k}^T \,) \ , \tag{2.4}$$

where $j, k = 0, 1, \cdots, p$.

In numerical computations with real data, calculating principal components (PCs) of $C_{t,j,k}$ with individual sample covariance matrixes directly one by one is not desirable due to limited size of sample. Instead, it is preferable to get PCs of $C_{t,j,k}$ indirectly by "dynamic, or spectral, principal components analysis (DPCA or SPCA) in frequency domain on Fourier transforms of the covariance matrixes $C_{t,j,k}$. Barigozzi (2024) and Doz and Fuleky (2020) reviewed, for example, estimations of stationary DFMs by DPCAs or SPCAs.

By extending the concept of DPCA/SPCA, the Appendixes A.1 through A.4 of this paper introduces, step-by-step, a non-stationary DFM estimation with following techniques:

- "conjugate two-dimensional discrete Fourier transform" (or conjugate 2D-DFT or C2D-DFT),
- spectral principal components analysis (or spectral PCA or SPCA) in frequency domain,
- inverse C2D-DFT,
- additional PCA in time domain, and
- classic (one-dimensional) DFT and inverse DFT.

This method does not involve any iterations (outside the publicly available PCA algorithms).

3. DFM Equations

Assume that we have computed principal components (PCs) of covariance matrixes $C_{t,j,k} = Cov(\, y_{t-j} \,,\, y_{t-k} \,)$ in Eq. (2.4) by the method of spectral PCA with conjugate 2D-DFT outlined above and detailed in Appendixes A.1 through A.4 of this paper. Then, based on the approaches discussed by Forni, Hallin, Lippi and Reichlin (2000), the time-varying coefficients and values, $X_{t,j}$, f_t and $A_{t,k}$ in DFM equations Eqs. (2.2) and (2.3) are estimated with the set of principal components of $C_{t,j,k}$, $j, k = 0, 1, \cdots, p$.

A dynamic factor model can be expressed in several different forms, with all forms derived from the same set of principal components. Different forms of DFM serve different applications of DFM. The measurement and state equations Eqs. (2.2) and (2.3) indicate that elements in factor loadings $X_{t,j}$ and factor VAR coefficients $A_{t,k}$ are functions of parameter set (t, j, k, p_1, p_2). Therefore, we will differentiate various forms of DFM by denoting the relevant matrixes as $X_{t,j}^{(p_1)}$ and $A_{t,k}^{(p_2)}$, respectively.

The first form of DFM of order p is of a single equation, with $p_1 = p$, as

$$y_t = X_{t,0}^{(p)} f_t + X_{t,1}^{(p)} f_{t-1} + \cdots + X_{t,p}^{(p)} f_{t-p} + u_t \,, \tag{3.1}$$

where Eq. (3.1) is in the same form as Eq. (1.4), y_t is $n \times 1$ <u>observed</u> vector representing n $(n \gg 1)$ individual time-series, f_t is $m \times 1$ <u>estimated</u> vector representing m $(m \ll n)$ common factors, $X_{t,j}^{(p)}$, $j = 0, 1, \cdots, p$, are $n \times m$ <u>estimated</u> matrixes of dynamic factor loadings, and u_t is $n \times 1$ <u>balancing</u> vector of stochastic idiosyncratic components. Comparing Eq. (3.1) with Eq. (2.2), we can call this form as "measurement equation only" form of DFM.

The second form of DFM is of a pair of equations, with $p_1 = 0$ and $p_2 = p$, as

$$y_t = X_{t,0}^{(0)} f_t + u_t \,, \tag{3.2}$$

$$f_t = A_{t,1}^{(p)} f_{t-1} + A_{t,2}^{(p)} f_{t-2} + \cdots + A_{t,p}^{(p)} f_{t-p} + v_t \,, \tag{3.3}$$

where Eqs. (3.2) and (3.3) are measurement and state equations, respectively, the <u>estimated</u> loadings matrix $X_{t,0}^{(0)}$ in Eq. (3.2) is different from the first loadings matrix $X_{t,0}^{(p)}$ in Eq. (3.1), $A_{t,k}^{(p)}$, $k = 1, 2, \cdots, p$, are $m \times m$ <u>estimated</u> matrixes of vector autoregressive (VAR) coefficients of common factors, and v_t is $m \times 1$ <u>unobserved</u> vector of stochastic factor innovations or prediction errors. Comparing Eq. (3.2) with Eq. (1.2), we can call this form as "static loadings" form of DFM, with all dynamics being modeled by the state equation Eq. (3.3).

The third form of DFM is a pair of equations, with $p_1 \geq 1$, $p_2 \geq 1$, $p_1 + p_2 = p$, and $p \geq 2$, as

$$y_t = X_{t,0}^{(p_1)} f_t + X_{t,1}^{(p_1)} f_{t-1} + \cdots + X_{t,p_1}^{(p_1)} f_{t-p_1} + u_t \ , \tag{3.4}$$

$$f_t = A_{t,1}^{(p_2)} f_{t-1} + A_{t,2}^{(p_2)} f_{t-2} + \cdots + A_{t,p_2}^{(p_2)} f_{t-p_2} + v_t \ , \tag{3.5}$$

where Eqs. (3.4) and (3.5) are measurement and state equations, respectively, and they are in the same pair of forms as Eqs. (2.2) and (2.3). We can call this form as "general" form of DFM, and values of p_1 and p_2 depend on the need(s) of specific application(s).

The common factor scores f_t participate in all DFM measurement and state equations Eqs. (3.1) through (3.5), where the values of f_t are estimated by the DFM modeling process. As an alternative method, factor scores f_t can be estimated with current and past values of observed vector time-series y_t as

$$f_t = S_{t,0}\, y_t + S_{t,1}\, y_{t-1} + S_{t,2}\, y_{t-2} + \cdots + S_{t,p}\, y_{t-p} + \xi_t \ , \tag{3.6}$$

where Eq. (3.6) can be referred to as a "common factor score filter" equation, $S_{t,j}$, $j = 0, 1, \cdots, p$, are $m \times n$ <u>estimated</u> matrixes of factor score filter coefficients, and ξ_t is $m \times 1$ unobserved vector of estimation errors or discrepancies between the filter Eq. (3.6) and other factor score estimation approaches. The filter coefficients $S_{t,j}$ are formulated based on the same set of principal components being used to estimate the factor loadings matrix $X_{t,j}^{(p_1)}$ and factor VAR coefficients $A_{t,k}^{(p_2)}$.

Based on similarity in format of expressions, we can pair up the factor score filter equation Eq. (3.6) and the time-series VAR equation Eq. (2.1), then compare this somewhat odd pair with a pair of measurement and state equations Eqs. (3.1) and (3.3). These two pairs mirror each other in format by switching between observed data y_t and unobserved factors f_t , but the two pairs of maximum possible ranks of coefficient matrixes in correspondent equations are different: they are (m, n) vs. (m, m) . By applying the set of principal components in our

DFM modeling to Eq. (2.1), it becomes a dimension-reduced VAR equation of observed vector time-series y_t as

$$y_t = M_{t,1}^{(DFM)}\, y_{t-1} + M_{t,2}^{(DFM)}\, y_{t-2} + \cdots + M_{t,p}^{(DFM)}\, y_{t-p} + e_t \,, \tag{3.7}$$

$$M_{t,k}^{(DFM)} = M^{(L)}\, M_{t,k}^{(R)} \,, \tag{3.8}$$

where $M^{(L)}$ is $n \times m$ estimated matrix serving as "left multiplier", $M_{t,k}^{(R)}$, $k = 1, 2, \cdots, p$, are $m \times n$ estimated matrixes serving as "right multipliers", $M_{t,k}^{(DFM)}$ are $n \times n$ matrixes with maximum possible rank of m, $m < n$ or $m \ll n$, and e_t is $n \times 1$ balancing vector of stochastic idiosyncratic components (resulting from both stochastic u_t and v_t components in measurement and state equations of DFM). Equation Eq. (3.7) is of "DFM-based" vector autoregression (VAR) of observed vector time-series y_t with dimension-reduced VAR coefficient matrixes $M_{t,k}^{(DFM)}$ by Eq. (3.8).

This paper is about frequentist (non-Bayesian) DFM estimations. Unlike Bayesian analysis, frequentist modeling does not treat all unobserved variables as stochastic. In frequentist DFM formulation, for example, factor loadings $X_{t,j}^{(p_1)}$, factor VAR coefficients $A_{t,k}^{(p_2)}$, time-series VAR coefficients $M_{t,k}^{(DFM)}$, and factor score filter coefficients $S_{t,j}$, are all estimated as deterministic. Only common factor scores f_t and random components u_t, v_t, ξ_t and e_t are treated as stochastic.

In our DFM estimation by spectral PCA with C2D-DFT, variances and vector autocovariances of common factor scores f_t,

$$V_{t-j} = Var(\, f_{t-j}\,) = E(\, f_{t-j}\, f_{t-j}^T\,) \,, \tag{3.9}$$

$$V_{(t-j),(t-k)} = Cov(\, f_{t-j}\,,\, f_{t-k}\,) = E(\, f_{t-j}\, f_{t-k}^T\,) \,, \tag{3.10}$$

are estimated by the same set of principal components of covariance matrixes $C_{t,j,k}$, $j, k =$

$0, 1, \ldots, p$. Based on commonly practiced DFM modeling assumptions, V_{t-j} is $m \times m$ diagonal matrix of variances, $V_{(t-j),(t-k)}$, $k \neq j$, is $m \times m$ matrix of vector autocovariances, random components u_t and v_t are not correlated with each other and not with factor scores f_{t-j} , and their variances, $R_t^{(u)} = Var(u_t)$ and $R_t^{(v)} = Var(v_t)$, are diagonal matrixes.

Derived from the measurement equation Eq. (3.4) of general form of DFM, covariance matrix of observed vector time-series y_t can be expressed as

$$\begin{aligned} C_t &= Var(y_t) = Var(\textstyle\sum_{j=0}^{p_1} X_{t,j}^{(p_1)} f_{t-j}) + Var(u_t) \\ &= \textstyle\sum_{j=0}^{p_1} \sum_{k=0}^{p_1} X_{t,j}^{(p_1)} Cov(f_{t-j}, f_{t-k}) (X_{t,k}^{(p_1)})^T + Var(u_t) \\ &= \textstyle\sum_{j=0}^{p_1} \sum_{k=0}^{p_1} [\, X_{t,j}^{(p_1)} V_{(t-j),(t-k)} (X_{t,k}^{(p_1)})^T \,] + R_t^{(u)} , \end{aligned} \tag{3.11}$$

where $V_{(t-j),(t-k)}$ is $m \times m$ <u>estimated</u> matrix of variance of factor f_{t-j} or vector autocovariance of factors f_{t-j} and f_{t-k} , and $R_t^{(u)}$ is $n \times n$ <u>estimated</u> diagonal matrix of variance of idiosyncratic components u_t .

Meanwhile, derived from the state equation Eq. (3.5) of general form of DFM, dynamic process of factor covariance matrix $V_{(t-j),(t-k)}$ in Eq. (3.11) above can be described by two equations. The first one is

$$\begin{aligned} V_t &= Var(f_t) = Var(\textstyle\sum_{j=1}^{p_2} A_{t,j}^{(p_2)} f_{t-j}) + Var(v_t) \\ &= \textstyle\sum_{j=1}^{p_2} \sum_{k=1}^{p_2} [\, A_{t,j}^{(p_2)} Cov(f_{t-j}, f_{t-k}) (A_{t,k}^{(p_2)})^T \,] + Var(v_t) \\ &= \textstyle\sum_{j=1}^{p_2} \sum_{k=1}^{p_2} [\, A_{t,j}^{(p_2)} V_{(t-j),(t-k)} (A_{t,k}^{(p_2)})^T \,] + R_t^{(v)} , \end{aligned} \tag{3.12}$$

where V_t is $m \times m$ diagonal matrix of variance of factor f_t , and $R_t^{(v)}$ is $m \times m$ <u>estimated</u> diagonal matrix of variance of innovation or prediction error v_t . The second one is

$$V_{t,(t-k)} = Cov(f_t, f_{t-k}) = Cov(\sum_{j=1}^{p_2} A_{t,j}^{(p_2)} f_{t-j}, f_{t-k})$$
$$= \sum_{j=1}^{p_2} A_{t,j}^{(p_2)} Cov(f_{t-j}, f_{t-k})$$
$$= \sum_{j=1}^{p_2} A_{t,j}^{(p_2)} V_{(t-j),(t-k)}, \qquad (3.13)$$

where $k = 1, 2, \ldots$, and $V_{t,(t-k)}$ is $m \times m$ vector autocovariance matrix of factors f_t and f_{t-k}. Since matrixes $A_{t,k}^{(p_2)}$ are vector autoregressive (VAR) coefficients of factor scores f_t, the variance Eq. (3.12) is a "quadratic VAR" equation, while the autocovariance Eq. (3.13) is a "linear VAR" equation.

In Eqs. (3.11), (3.12) and (3.13), there are $p_1 + 1$ of $n \times m$ factor loadings matrixes $X_{t,j}^{(p_1)}$ and p_2 of $m \times m$ factor VAR matrixes $A_{t,k}^{(p_2)}$. In big data analysis where $n \gg m$, a loadings matrix $X_{t,j}^{(p_1)}$ is much larger than a factor VAR matrix $A_{t,k}^{(p_2)}$.

When taking measurement and state equations Eqs. (3.2) and (3.3) of static loadings form of DFM, the above three factor covariance equations become

$$C_t = Var(y_t) = Var(X_{t,0}^{(0)} f_t) + Var(u_t)$$
$$= X_{t,0}^{(0)} V_t (X_{t,0}^{(0)})^T + R_t^{(u)}, \qquad (3.14)$$

$$V_t = Var(f_t) = Var(\sum_{j=1}^{p} A_{t,j}^{(p)} f_{t-j}) + Var(v_t)$$
$$= \sum_{j=1}^{p} \sum_{k=1}^{p} [A_{t,j}^{(p)} V_{(t-j),(t-k)} (A_{t,k}^{(p)})^T] + R_t^{(v)}, \qquad (3.15)$$

$$V_{t,(t-k)} = Cov(f_t, f_{t-k}) = Cov(\sum_{j=1}^{p} A_{t,j}^{(p)} f_{t-j}, f_{t-k})$$
$$= \sum_{j=1}^{p} A_{t,j}^{(p)} V_{(t-j),(t-k)}. \qquad (3.16)$$

In Eqs. (3.14), (3.15) and (3.16), there is only one $n \times m$ factor loadings matrix $X_{t,0}^{(0)}$ and p of $m \times m$ factor VAR matrixes $A_{t,k}^{(p)}$. Therefore, the three equations above (based on the static loadings form of DFM) are simpler than those Eqs. (3.11), (3.12) and (3.13) (based on the general form of DFM), and smaller (in terms of total number loadings and VAR coefficients) as well when modeling big datasets where $n \gg m$.

As summarized at the end of Appendix A.4, in a non-stationary dynamic factor model estimation,

- dynamic factor loadings matrixes $X_{t,j}^{(p_1)}$,
- common factor scores f_t,
- dynamic factor VAR matrixes $A_{t,k}^{(p_2)}$,
- variances V_{t-j} of factor scores,
- vector autocovariances $V_{(t-j),(t-k)}$ of factor scores,
- factors score filter coefficient matrixes $S_{t,j}$, and
- DFM-based VAR matrixes $M^{(L)}$ and $M_{t,k}^{(R)}$

can be estimated by results of spectral PCA with conjugate 2D-DFT:

- principal eigenvectors $H_{t,j}^{(C)}$ in Eq. (A.4.9),
- principal eigen scores $f_{t-j}^{(C)}$ by Eq. (A.4.14),
- principal eigenvalues $P_{t,j}^{(C)} = P_{t,j,j}^{(C)}$ in Eq. (A.4.9) and
- (principal) coefficient matrixes $P_{t,j,k}^{(C)}$ by Eq. (A.4.11),

where additional PCA in time domain and classic one-dimensional DFT are also involved.

4. Vector Time-Series Forecasts

State equations of DFMs, e.g. Eq. (3.3) or (3.5), are of vector autoregressions (VAR). This fact indicates that a primary application of dynamic factor models is to forecast future values of observed vector time-series y_t.

To conform with common conventions, we denote s-step forecasts of data y_t given all current and past observed data available at time t as $y_{(t+s)|t}$. We will denote s-step forecasts of unobserved factors f_t and random components u_t and v_t in the same way. By replacing time index subscript t with $(t+s)|t$, the state equation of general form of DFM, Eq. (3.5), becomes

$$f_{(t+s)|t} = \sum_{k=1}^{p_2} A_{t,k}^{(p_2)} f_{(t+s-k)|t} + v_{(t+s)|t} \ ,$$

where forecasting step $s = 1, 2, \cdots$, while the measurement equation of general form of DFM, Eq. (3.4), becomes

$$y_{(t+s)|t} = \sum_{j=0}^{p_1} X_{t,j}^{(p_1)} f_{(t+s-j)|t} + u_{(t+s)|t} \ .$$

This pair of equations demonstrates that, with an estimated dynamic factor model, we can make stepwise forecasts of unobserved common factors f_t first, and then make multi-step forecasts of observed vector time-series y_t. The values of factors $f_{(t+s-j)|t}$ are <u>forecasted</u> when $(t+s-j) > t$, or <u>estimated</u> f_{t+s-j} when $(t+s-j) \leq t$. The estimated factor scores f_{t+s-j} can serve as anchors in stepwise forecasts. To avoid an awkward case of forecasting only by earlier forecasts without any anchoring, we limit the forecasting steps by setting, for example,

$$1 \leq s \leq s_{max} < p_2 = p - p_1 \ .$$

When the random components u_t and v_t are not further modeled, we will assume:

$$u_{(t+s)|t} = E\left(u_{(t+s)|t} \right) = E(u_t) = 0 \ ,$$

$$v_{(t+s)|t} = E\left(v_{(t+s)|t} \right) = E(v_t) = 0 \ .$$

Adding up all discussions above, we can make stepwise forecasts of unobserved common factors f_t as

$$f_{(t+s)|t} = \sum_{k=1}^{p_2} A_{t,k}^{(p_2)} f_{(t+s-k)|t} \ , \tag{4.1}$$

and then make multi-step forecasts of observed vector time-series y_t as

$$y_{(t+s)|t} = \sum_{j=0}^{p_1} X_{t,j}^{(p_1)} f_{(t+s-j)|t} \ , \tag{4.2}$$

where forecasting step $s = 1, 2, \cdots, s_{max} < p_2 = p - p_1$ in both Eqs. (4.1) and (4.2), and $f_{(t+s-j)|t} = f_{t+s-j}$ when $(t + s - j) \leq t$.

When a larger maximum number of forecasting step s_{max} is expected, we can use the static loadings form of DFM, i.e. Eqs. (3.2) and (3.3), where $p_1 = 0$ and $p_2 = p$, to make factor forecasts as

$$f_{(t+s)|t} = \sum_{k=1}^{p} A_{t,k}^{(p)} f_{(t+s-k)|t} \ , \tag{4.3}$$

and then vector time-series forecasts as

$$y_{(t+s)|t} = X_{t,0}^{(0)} f_{(t+s)|t} \ , \tag{4.4}$$

where forecasting step $s = 1, 2, \cdots, s_{max} < p$ in both Eqs. (4.3) and (4.4), and Eq. (4.4) is of a single "static loadings" matrix $X_{t,0}^{(0)}$.

When making forecasts by non-stationary (time-varying) DFM, it would be ideal to forecast both VAR coefficients $A_{t,k}^{(p_2)}$ and loadings matrixes $X_{t,j}^{(p_1)}$ as well. In DFM modeling discussed in this paper, however, they are not forecasted, but simply assumed "estimates as forecasts":

$$A_{[(t+s)|t],k}^{(p_2)} = A_{t,k}^{(p_2)} \ , \tag{4.5}$$

$$X_{[(t+s)|t],j}^{(p_1)} = X_{t,j}^{(p_1)} \ , \tag{4.6}$$

where $0 \leq p_1 < p$, $1 \leq p_2 \leq p$, and $p_1 + p_2 = p$. Errors caused by such estimates as forecasts will be added to the overall prediction errors.

In addition, the DFM-based dimension-reduced vector autoregressive equation of observed data y_t , Eqs. (3.7) and (3.8), indicate that vector time-series y_t can be forecasted directly by Eq. (3.7) without explicit forecasts of factor scores f_t . By comparing time-series

VAR Eq. (3.7) with factor VAR Eq. (3.3), and assuming $e_{(t+s)|t} = E(\, e_{t+s}\,) = E(\, e_t\,) = 0$ (same as "forecasting" u_t and v_t), we can make stepwise forecasts of vector time-series y_t as

$$y_{(t+s)|t} = M^{(L)} \, [\, \sum_{j=1}^{p} M_{t,j}^{(R)} \, y_{(t+s-j)|t} \,] \, , \tag{4.7}$$

where forecasting step $s = 1, 2, \cdots, s_{max} < p$, values of vector time-series $y_{(t+s-j)|t}$ are forecasted values when $(t + s - j) > t$ and are actually observed data y_{t+s-j} when $(t + s - j) \leq t$, $M^{(L)}$ is $n \times m$ matrix, and $M_{t,k}^{(R)}$ are $m \times n$ matrixes.

5. Shock Response Forecasts

A dynamic factor model (DFM) estimated on a (large) set of observed time-series describes a specific vector autoregressive (VAR) process which can explain the training dataset. When this VAR process takes some value(s) as inputs, its outputs are specific forecasts given the modeled VAR process and the inputs. The previous Section 4 discussed DFM forecasts when taking observed vector time-series as inputs. This section discusses DFM forecasts (or responses) when taking exogenous shocks (or impulses) as inputs.

Denote a set of simultaneous shocks on vector time-series y_t by a vector $y^{(Shock)}$, of which some elements are positive, some negative, while all remaining others are zeros. Assume that the shocks occurred at time $t - j$ as

$$y_{t-j}^{(Shock)} = y^{(Shock)} \, , \tag{5.1}$$

where $0 \leq j \leq p$ (most likely $j = 0$), and

$$y_{t-\tau}^{(Shock)} = 0 \quad \forall \quad \tau \geq 0 \;\; and \;\; \tau \neq j \, , \tag{5.2}$$

i.e. all shocks occurred at a single moment in time.

To make shock response forecasts using a state space representation of DFM, we need to estimate shock-induced factor score time-series first. Substituting the shock vector time-series $y_t = y_t^{(Shock)}$ into dynamic factor score filter equation Eq. (3.6) yields shock-induced factor score time-series as

$$f_t^{(Shock)} = \sum_{k=0}^{p} S_{t,k}\, y_{t-k}^{(Shock)} , \tag{5.3}$$

where $S_{t,k}$ are $m \times n$ matrixes of factor score filter coefficients. Substituting factors $f_t = f_t^{(Shock)}$ into DFM forecast equations, Eqs. (4.1) and (4.2) or Eqs. (4.3) and (4.4), and denoting

$$f_{t+s-k} = f_{t+s-k}^{(Shock)} \quad \forall \quad k \geq s , \tag{5.4}$$

$$y_{t+s-k} = y_{t+s-k}^{(Shock)} \quad \forall \quad k \geq s , \tag{5.5}$$

the forecasted shock responses, $f_{(t+s)|t}^{(Resp)}$ and $y_{(t+s)|t}^{(Resp)}$ are

$$f_{(t+s)|t}^{(Resp)} = f_{(t+s)|t} , \tag{5.6}$$

$$y_{(t+s)|t}^{(Resp)} = y_{(t+s)|t} , \tag{5.7}$$

where forecasting step $s = 1, 2, \cdots, s_{max}$, $f_{(t+s)|t}$ are factor score forecasts made by Eq. (4.1) or (4.3), and $y_{(t+s)|t}$ are vector time-series forecasts made by Eq. (4.2) or (4.4),

To make shock response forecasts using DFM-based vector autoregressive (VAR) equation Eq. (4.7), substituting shock vector time-series $y_t = y_t^{(Shock)}$ into Eq. (4.7), we can get forecasted shock response $y_{(t+s)|t}^{(Resp)} = y_{(t+s)|t}$ directly.

6. Multivariate Volatility Forecasts

The fact that DFMs are estimated on observed, not squared, values of vector time-series might imply that they may not be effective for forecasting variances and covariances of vector time-series. As comparison, models, e.g. various GARCH models, estimated explicitly on squared values of time-series have become popular models for forecasting variances of time-series. When DFM time-series forecasts are good, squared values of the forecasts could be good predictors as well.

In fact, the quadratic VAR and linear VAR equations, Eqs. (3.15) and (3.16) or Eqs. (3.12) and (3.13), indicate that DFMs can make effective forecasts of variances and covariances of observed vector time-series. As a result, DFMs can forecast both values and covariances of multiple time-series by the same models estimated on the same training datasets.

Let's consider the simpler and smaller form of covariance expressions Eqs. (3.14), (3.15) and (3.16) at first. As in Section 4, we denote s-step forecasts of covariance matrix C_t (of vector time-series y_t) given all current and past observed data available at time t as $C_{(t+s)|t}$. We will denote s-step forecasts of factor variances V_t and error variances $R_t^{(u)}$ and $R_t^{(v)}$ in the same way. By replacing time index subscript t with $(t+s)|t$, the factor variance equation Eq. (3.15) becomes

$$
\begin{aligned}
&V_{(t+s)|t} = Var(\, f_{(t+s)|t}\,) = Var(\, \textstyle\sum_{j=1}^{p} A_{t,j}^{(p)} f_{(t+s-j)|t}\,) + Var(\, v_{t+s)|t}\,) \\
&= \textstyle\sum_{j=1}^{p} \sum_{k=1}^{p} [\, A_{t,j}^{(p)}\, V_{(t+s-j),(t+s-k)|t}\, (A_{t,k}^{(p)})^T\,] + R_{(t+s)|t}^{(v)}\,, \qquad (6.1)
\end{aligned}
$$

where forecasting step $s = 1, 2, \cdots,$ the factor autocovariance equation Eq. (3.16) becomes

$$
\begin{aligned}
&V_{(t+s),(t+s-k)|t} \\
&= Cov(\, f_{(t+s)|t}, f_{(t+s-k)|t}\,) = Cov(\, \textstyle\sum_{j=1}^{p} A_{t,j}^{(p)} f_{(t+s-j)|t}\,, f_{(t+s-k)|t}\,) \\
&= \textstyle\sum_{j=1}^{p} A_{t,j}^{(p)}\, V_{(t+s-j),(t+s-k)|t}\,, \qquad (6.2)
\end{aligned}
$$

while the covariance equation Eq. (3.14) becomes

$$
\begin{aligned}
&C_{(t+s)|t} = Var(\, y_{(t+s)|t}\,) = Var(\, X_{t,0}^{(0)}\, f_{(t+s)|t}\,) + Var(\, u_{(t+s)|t}\,) \\
&= X_{t,0}^{(0)}\, V_{(t+s)|t}\, (X_{t,0}^{(0)})^T + R_{(t+s)|t}^{(u)}\,.
\end{aligned}
$$

These three equations demonstrate that, with an estimated DFM, we can make stepwise forecasts of factor variances V_t first, and then make multi-step forecasts of covariance C_t of observed vector time-series y_t. The values of factor covariances $V_{(t+s-j),(t+s-k)|t}$ are forecasted when $(t+s-j) > t$ or $(t+s-k) > t$, or estimated $V_{(t+s-j),(t+s-k)}$ when both $(t+s-j) \leq t$ and $(t+s-k) \leq t$. The estimated factor covariances $V_{(t+s-j),(t+s-k)}$ can serve as anchors in stepwise forecasts. To avoid an awkward case of forecasting only by earlier forecasts without any anchoring, we limit the forecasting steps by setting, for example,

$$1 \leq s \leq s_{max} < p \,.$$

When the random components u_t and v_t are not further modeled, we will assume:

$$R_{(t+s)|t}^{(u)} = R_{t+s}^{(u)} = R_t^{(u)} \,,$$

$$R_{(t+s)|t}^{(v)} = R_{t+s}^{(v)} = R_t^{(v)} \,.$$

In estimation of non-stationary DFM by spectral PCA with C2D-DFT and additional PCA in time domain, symmetric covariance matrix C_t of observed data vector y_t is diagonalized and dimension-reduced by time-varying principal eigenpairs (eigenvalues and eigenvectors). As a result, variance matrix V_t of unobserved factors f_t is diagonal (with diagonal elements equal to principal eigenvalues). It would be ideal to forecast principal eigenpairs as well. In DFM modeling discussed in this paper, however, they are not forecasted, but simply assumed “estimates as forecasts”, same as non-stationary DFM coefficient matrixes $A_{t,k}^{(p_2)}$ and $X_{t,j}^{(p_1)}$ in Eqs. (4.5), (4.6), (6.1) and (6.2). Therefore, forecasted factor variance $V_{(t+s)|t}$ by Eqs. (6.1) with (6.2) is generally not diagonal. Taking only diagonal elements of $V_{(t+s)|t}$ and setting all off-diagonal ones as zeros is to be consistent with DFM modeling assumptions.

Adding up all discussions above, we can make stepwise forecasts of factor variance matrix V_t as

$$V_{(t+s)|t}$$
$$= \sum_{j=1}^{p} \sum_{k=1}^{p} Diag(A_{t,j}^{(p)} V_{(t+s-j),(t+s-k)|t} (A_{t,k}^{(p)})^T) + R_t^{(v)} , \qquad (6.3)$$

where function $Diag(V)$ converts square matrix V (symmetric matrix, in this case) into diagonal matrix by setting all off-diagonal elements of V as 0. And make stepwise forecasts of factor autocovariance matrix $V_{t,(t-k)}$ as

$$V_{(t+s),(t+s-k)|t} = \sum_{j=1}^{p} A_{t,j}^{(p)} V_{(t+s-j),(t+s-k)|t} . \qquad (6.4)$$

Then make multi-step forecast of covariance matrix C_t as

$$C_{(t+s)|t} = X_{t,0}^{(0)} V_{(t+s)|t} (X_{t,0}^{(0)})^T + R_t^{(u)} . \qquad (6.5)$$

Forecasting step $s = 1, 2, \cdots, s_{max} < p$ in all three equations Eqs. (6.3), (6.4) and (6.5), and autoregression time-lag $k = 1, 2, \cdots$ in Eq. (6.4).

Multivariate volatility forecasts by Eqs. (6.3), (6.4) and (6.5) are derived from static loadings form of DFM by Eqs. (3.2) and (3.3) where $p_1 = 0$ and $p_2 = p$. When starting with a general form of DFM by Eqs. (3.4) and (3.5) where both $p_1 \geq 1$ and $p_2 \geq 1$ hold, we can make stepwise forecasts of factor variance matrix V_t as

$$V_{(t+s)|t}$$
$$= \sum_{j=1}^{p_2} \sum_{k=1}^{p_2} Diag(A_{t,j}^{(p_2)} V_{(t+s-j),(t+s-k)|t} (A_{t,k}^{(p_2)})^T) + R_t^{(v)} , \qquad (6.6)$$

and make stepwise forecasts of factor autocovariance matrix $V_{t,(t-k)}$ as

$$V_{(t+s),(t+s-k)|t} = \sum_{j=1}^{p_2} A_{t,j}^{(p_2)} V_{(t+s-j),(t+s-k)|t} , \qquad (6.7)$$

then make multi-step forecast of covariance matrix C_t as

$$C_{(t+s)|t}$$
$$= \sum_{j=0}^{p_1} \sum_{k=0}^{p_1} [X_{t,j}^{(p_1)} V_{(t+s-j),(t+s-k)|t} (X_{t,k}^{(p_1)})^T] + R_t^{(u)} , \qquad (6.8)$$

where forecasting step $s = 1, 2, \cdots, s_{max} < p_2 < p$.

7. Performance Metrics and Benchmarks

Expressions of time index subscript in time-series forecasts Eqs (4.1) and (4.2) and volatility forecasts Eqs. (6.3), (6.4) and (6.5) suggest chronological splits of observed vector time-series data into three categories of datasets: (1) rolling training datasets in moving windows for rolling model estimations, (2) a validation dataset for hyperparameter selection, and (3) a test dataset for final performance evaluation of forecasts.

Training a dynamic factor model is to estimate coefficients and values in factor loadings matrixes $X_{t,j}^{(p_1)}$, factor VAR matrixes $A_{t,k}^{(p_2)}$, factor covariance matrixes $V_{(t-j),(t-k)}$, and diagonal variance matrixes $R_t^{(u)}$ and $R_t^{(v)}$. Hyperparameters include DFM orders (i.e. ranges of time lags) p , p_1 and p_2 , number of common factors m , and size of rolling (or moving) training data windows. There has been extensive research on how to determine correct values of model orders and number of common factors, which is not discussed in this paper. In practice, hyperparameter tuning is to search for optimal values of hyperparameters. With datasets quantifying real world events, where unexpected shocks and independent noisy activities making stochastic components strong or even dominant, signal to noise ratios of forecasts are often quite low and/or need to be measured carefully using different performance metrics.

Assume, in our DFM analysis, time-index t of observed vector time-series y_t covers an entire time period $t = 0, 1, \cdots, T$, model order is p , size of time window for test dataset is T_{test} , size for validation dataset is T_{valid} , size of rolling (or moving) window for training datasets is T_{train} , maximum forecasting step is s_{max} , and total length of vector time-series $T \geq (T_{train} + T_{valid} + T_{test} + p + 2\, s_{max})$.

The test dataset is of data points y_t at $t = t_{test}$ in the last segment of the observed vector time-series,

$$(T - T_{test} - s_{max}) < t_{test} \leq T \ , \tag{7.1}$$

and the validation dataset is of data points y_t at $t = t_{valid}$ in the last segment earlier than the test dataset,

$$(T - T_{test} - T_{valid} - 2\ s_{max}) < t_{valid} \leq (T - T_{test} - s_{max}) \ . \tag{7.2}$$

For all observed data points y_t at $t = t_{eval}$ in either test or validation dataset,

$$(T - T_{test} - T_{valid} - 2\ s_{max}) < t_{eval} \leq T \ , \tag{7.3}$$

we will forecast each of them with all candidate DFM models. Each model is estimated by the most recent observations y_t at $t = t_{train}$ in the training dataset, that must be prior to making s-step forecast, i.e. $(t_{train} + s) \leq t_{eval}$, or

$$(t_{eval} - s - T_{train} - p) < t_{train} \leq (t_{eval} - s) \ . \tag{7.4}$$

where time $t = (t_{eval} - s)$ is when the model makes s-step forecasts of data point y_τ at $\tau = t + s = t_{eval}$.

In evaluation of time-series forecasts, comparing forecasted values $y_{(t+s)|t}$ vs. available observed values y_τ at $\tau = t + s = t_{eval}$ using appropriate metrics, performances of forecasts can be objectively measured or scored. Some popular metrics provided by publicly available machine learning packages, such as scikit-learn (https://scikit-learn.org), can be used to score DFM time-series forecasts. Metrics measuring similarity or dissimilarity between forecasted and observed values available in scikit-learn (https://scikit-learn.org/stable/modules/model_evaluation.html) include:

- d2_absolute_error_score,
- explained_variance_score,

- max_error,
- mean_absolute_error,
- mean_absolute_percentage_error,
- mean_squared_error,
- mean_squared_log_error,
- median_absolute_error, and
- r2_score.

Other useful metrics can be coded easily include:

- Pearson correlation coefficient between forecasted values and observed data in validation or test dataset, and
- high-dimensional projection coefficient of forecasted values onto observed data space, with size of validation or test dataset to be used as number of dimensions of projection.

Since realized covariance matrix C_t of vector time-series y_t is not directly observable, evaluating forecasted covariance matrix $C_{(t+s)|t}$ would be more involved. A widely practiced evaluation technique is to measure accuracy of forecasted variance of a weighted sum of time-series

$$y_t^{(w)} = w^T y_t = y_t^T w, \tag{7.5}$$

where w is a $n \times 1$ vector of weights for aggregation. Forecasted variance of aggregated time-series $y_t^{(w)}$ is

$$Var(y_{(t+s)|t}^{(w)}) = (\sigma_{(t+s)|t}^{(w)})^2 = w^T C_{(t+s)|t} w . \tag{7.6}$$

To evaluate forecasted variance of an individual time-series, we can set an element of w as 1 and all others as 0. To evaluate forecasted covariance of a pair of two time-series, we can set two elements of w as 0.5 each and all others as 0. This way, all elements of forecasted covariance matrix $C_{(t+s)|t}$ can be evaluated one by one.

To measure the accuracy of forecasted variance $(\sigma^{(w)}_{(t+s)|t})^2$ in Eq. (7.6), a "realized z-score squared of the forecasts" defined by

$$(z^{(w)}_{(t+s)|t})^2 = (\, y^{(w)}_{t+s} - \mu^{(w)}_{t+s} \,)^2 \,/\, (\sigma^{(w)}_{(t+s)|t})^2 \,, \tag{7.7}$$

is handy. Observation $y^{(w)}_{t+s}$ and estimate $\mu^{(w)}_{t+s}$ are made at time $t + s = t_{eval}$, while forecast $(\sigma^{(w)}_{(t+s)|t})^2$ is made at earlier time t, and $\mu^{(w)}_{t+s} \equiv 0$ when vector time-series y_t are with mean values removed. According to a number of papers cited by i4cast LLC (2023), the accuracy of variance forecasts over validation or test dataset $(t + s) \in [t_1, t_2]$ can be measured by bias statistic $BS^{(w)}_s$, log-likelihood $LL^{(w)}_s$, and Q-statistic $QS^{(w)}_s$ defined as

$$BS^{(w)}_s = [\, \frac{1}{t_2 - t_1} \sum_{t=t_1-s}^{t_2-s} (z^{(w)}_{(t+s)|t})^2 \,]^{1/2} \,, \tag{7.8}$$

$$LL^{(w)}_s = -\, \frac{1/2}{t_2 - t_1} \sum_{t=t_1-s}^{t_2-s} [\, \ln(2\pi) + (z^{(w)}_{(t+s)|t})^2 + \ln(\sigma^{(w)}_{(t+s)|t})^2 \,] \,, \tag{7.9}$$

$$QS^{(w)}_s = \frac{1}{t_2 - t_1} \sum_{t=t_1-s}^{t_2-s} [\, (z^{(w)}_{(t+s)|t})^2 - \ln(z^{(w)}_{(t+s)|t})^2 \,] \,. \tag{7.10}$$

A bias statistic $BS^{(w)}_s > 1$ or $BS^{(w)}_s < 1$ shows an under- or over-prediction of volatility. A higher log-likelihood $LL^{(w)}_s$ or a lower Q-statistic $QS^{(w)}_s$ indicates more accurate forecasts.

All metrics discussed above measure similarity or dissimilarity between forecasted outcomes vs. observed data or information. The resulted performance scores of forecasts can be referred to as "absolute" scores. In many cases, "relative" performance scores of forecasts by models vs. specific simple forecasts can offer additional insights into the models.

Given a validation or test dataset of observed vector time-series data points, and given a forecasting step s, difference of performance score of s-step forecasts made by Model A minus score by Method B reflect value added or lost by Model A relative to Method B. The forecasts by Method B can be referred to as performance benchmark. "Naïve", or non-informational, simple forecasts are widely utilized as objective performance benchmarks in model evaluations. The best-known naïve benchmark for time-series forecast is the mean value of the time-series, which is 0 for zero-mean time-series. Most of the naïve benchmark forecasts are essentially mean values of pure random forecasts.

Therefore, relative scores of model forecasts vs. naïve benchmarks, instead of absolute scores themselves, are widely used in evaluation of forecasts by statistic models. Some data-driven models may not be any better than naïve benchmarks. Overfitted models, for example, mistaking noises as forecastable signals, are most often unable to outperform naïve benchmarks.

To evaluate variance/covariance forecasts, the simplest naïve forecast is "sample covariance as forecast":

$$C_t^{(Sample)} = K^{-1} \sum_{k=0}^{K} (\, y_{t-k} - \mu_t \,)\, (\, y_{t-k} - \mu_t \,)^T \,, \qquad (7.11)$$

$$C_{(t+s)|t}^{(Sample)} = C_t^{(Sample)} \,, \qquad (7.12)$$

where forecasting step $s = 1,2, \dots$. Another popular benchmark is "static factor model estimate as forecast":

$$C_t^{(Estimate)} = X_t \, V_t \, X_t^T + R_t \,, \qquad (7.13)$$

$$C_{(t+s)|t}^{(Estimate)} = C_t^{(Estimate)} \,, \qquad (7.14)$$

where $s = 1,2, \dots,$ and static factor loadings X_t, diagonal matrix of factor variance V_t, and diagonal matrix of residual variance R_t are estimated by a static factor model as:

$$y_t = \mu_t + X_t \, f_t + u_t$$

$$V_t = Var(\, f_t \,) = E(\, f_t \, f_t^T \,)$$

$$R_t = Var(\, u_t \,) = E(\, u_t \, u_t^T \,)$$

A comparison between the static benchmark Eq. (7.13) with our dynamic formula Eq. (6.5) indicates that a very handy benchmark is "nowcast serving as forecast":

$$C_t^{(Nowcast)} = X_{t,0}^{(0)} V_t \, (X_{t,0}^{(0)})^T + R_t^{(u)} \, , \tag{7.15}$$

$$C_{(t+s)|t}^{(Nowcast)} = C_t^{(Nowcast)} \, , \tag{7.16}$$

where benchmark Eq. (7.15) is a nowcast obtained by substituting $s = 0$ into non-stationary DFM equation Eq. (6.5).

8. Examples

"Long-Memory Dynamic Factor Model" algorithm (LMDFM, https://aws.amazon.com/marketplace/pp/prodview-da6ffrp4mlopg?sr=0) publicly available on AWS is an implementation of non-stationary dynamic factor model estimated by spectral PCA with conjugate 2D-DFT presented in this paper.

The data file included in the LMDFM algorithm package (https://github.com/i4cast/aws/blob/main/long_memory_dynamic_factor_model/input/Weekly_VTS_6Yr.csv) contains an example of vector time-series dataset. It is of weekly total returns in logarithm of tens of U.S. ETFs listed and traded in the U.S. market.

9. Discussion

This paper introduces an estimation of large non-stationary dynamic factor models (DFMs) by spectral PCA with conjugate 2D-DFT. An implementation of this estimation approach is shown as an example in Section 8 above. Time-series forecasts and variance/covariance forecasts made by this model fitted appropriately can outperform benchmark forecasts significantly.

This paper, however, does not model stochastic idiosyncratic components u_t in measurement equation Eq. (2.2), (3.2) or (3.4) of DFMs. Predictive powers of time-series forecasts and variance/covariance forecasts are therefore limited by such omissions.

The idiosyncratic components u_t can be modeled, analyzed and forecasted by "Yule-Walker PCA Autoregressive Model" algorithm (YWpcAR, https://aws.amazon.com/marketplace/pp/prodview-prndys7tr7go6?sr=0) available on AWS. We will discuss YWpcAR modeling in a separate paper.

To make better time-series forecasts of vector time-series, "Long-Memory Vector Autoregressive Model" (LMVAR, https://aws.amazon.com/marketplace/pp/prodview-335ruebzkb4bc?sr=0) available on AWS combines modeling and time-series forecasting functionalities of both LMDFM and YWpcAR algorithms.

To make better variance/covariance forecasts of vector time-series, "Dynamic Factor Variance-Covariance Model", (DFVCM, https://aws.amazon.com/marketplace/pp/prodview-yvaulquatt3v2?sr=0) available on AWS combines modeling and variance/covariance forecasting functionalities of both LMDFM and YWpcAR algorithms.

10. Further Development

Current implementations (e.g. LMDFM algorithm available on AWS as cited in Section 8, https://aws.amazon.com/marketplace/pp/prodview-da6ffrp4mlopg?sr=0) of estimating non-stationary DFMs by spectral PCA with conjugate 2D-DFT, as detailed in Appendixes A.1 through A.4 of this paper, are moving window estimations, as if stationarity holds within each training data window of limited length in time.

A variety of Bayesian filters are developed on vector time-series datasets, to make better non-stationary analyses and forecasts. A Bayesian filter is an algorithm applying Bayesian theorem recursively, updating unknown distributions of unobserved hidden variables as posterior, when current observations become available serving as data, where prior is the unobserved distributions estimated or updated based on all observations before the current ones became available. Classic Kalman filter is the best-known Bayesian filter. Other Bayesian filters are for different unobserved variables and are versatile. Particle filters are digital implementations of Bayesian filters and are computationally intensive and time-consuming. Variational Bayesian filters are analytic approximations of Bayesian filters and are among the fastest Bayesian filters.

Specifically, non-stationary covariance matrix $C_{t,j,k} = E(\, y_{t-j}\, y_{t-k}^T\,)$ of observed data vector y_t defined by Eq. (2.4) can be estimated by a Bayesian filter Eq. (A.1.3) as

$$C_{t,j,k}^{(BF)} = F(\, y_{t-j}\, y_{t-k}^T \mid C_{(t-1),j,k}^{(BF)} \,;\, \theta_C\,)\,, \tag{10.1}$$

where, of Bayesian analysis of the filter, $C_{t,j,k}^{(BF)}$ is posterior, y_t is current observations as data, $C_{(t-1),j,k}^{(BF)}$ serves as prior at time t, and θ_C is a set of all parameters (may include assumptions of distributions of y_t and $C_{t,j,k}$).

When applying classic one-dimensional discrete Fourier transform Eq. (A.1.10) on observed vector time-series y_t,

$$z_{t,m} = \sum_{j=0}^{p} y_{t-j}\, exp[-i2\pi mj \,/\, (p+1)]\,, \tag{10.2}$$

where $m = 0, 1, \dots, p$, covariance (or autocovariance) matrixes of the complex-value vector time-series $z_{t,m}$ and $z_{t,n}$ are as Eq. (A.1.12) as

$$W_{t,m,n} = Cov(\, z_{t,m}\,,\, z_{t,n}\,) = S_{t,m,n}\,,$$

where $m, n = 0, 1, \dots, p$. Spectral density matrixes, $S_{t,m,n}$, are actually covariance or autocovariance matrixes of frequency components $z_{t,m}$ and $z_{t,n}$ of vector time-series y_t at specific frequencies.

Similar to Eq. (10.1), the spectral density matrixes, $S_{t,m,n}$, can be directly estimated by a Bayesian filter Eq. (A.1.13) as

$$S_{t,m,n}^{(BF)} = F(\, z_{t,m}\, z_{t,n}^{H} \mid S_{(t-1),m,n}^{(BF)}\, ;\ \theta_S\,)\ , \tag{10.3}$$

where, of Bayesian analysis of the filter, $S_{t,m,n}^{(BF)}$ is posterior, $z_{t,m}$ is current information as data, $S_{(t-1),m,,n}^{(BF)}$ serves as prior at time t, and θ_S is a set of all parameters (may include assumptions of distributions of $z_{t,m}$ and $S_{t,m,n}$).

Developing a variational Bayesian filter (VBF) for Eq. (10.1) or Eq. (10.3) in our non-stationary DFM estimation by spectral PCA with conjugate 2D-DFT is our next step in our R&D efforts. VBFs are analytic approximations of Bayesian filters (as opposed to computational particle filters) and, therefore, best known for real-time, or on-line, learning and inference. A VBF-based DFM will be able to serve as an "on-line DFM" to be utilized for

- real-time analyses and forecasts of values and covariances of observed vector time-series,
- real-time tracking of multiple moving targets, and
- real-time industrial control process, etc.

Our team is an advanced developer of variational Bayesian filtering, demonstrated by Variational Bayesian Filtering Factor Analysis algorithm (VBfFA, https://aws.amazon.com/marketplace/pp/prodview-vdwcbntcsnu72?sr=0) available on AWS.

A.1. Conjugate 2D-DFT

Assume that we have a large number of stochastic time-series, collectively denoted as time-series of stochastic vector y_t, with each element of y_t representing an individual time-series. And assume that this large set of time-series can be described by a non-stationary, or time-varying, dynamic factor model (DFM). Appendixes A.1 through A.4 introduce, step-by-step, a non-stationary DFM estimation by "conjugate two-dimensional discrete Fourier transform" (conjugate 2D-DFT, or C2D-DFT), spectral principal components analysis (spectral PCA, or SPCA) in frequency domain, inverse C2D-DFT (Inv-C2D-DFT), additional PCA in time domain, and additional DFT and inverse DFT.

Given a zero-mean stochastic vector time-series y_t, an assumption that vector autoregression (VAR) of y_t is of order p, and a pair of time-lags j and k, a set of variance-covariance matrixes (VCMs) and vector autocovariance matrixes (ACMs) can be defined as:

$$C_{t,j,k} = Cov(\, y_{t-j}\, ,\, y_{t-k}\,) = E(\, y_{t-j}\, y_{t-k}^T\,)\,, \qquad \text{(A.1.1)}$$

$$C_{t,j,k} = C_{t,k,j}^T\,, \qquad \text{(A.1.2)}$$

where covariance matrix $C_{t,j,k}$ is VCM when $k = j$ or ACM when $k \neq j$, and $j, k = 0, 1, \ldots, p$.

There are several different ways to estimate covariance matrixes $C_{t,j,k}$. The simplest technique is "realized trailing sample covariance",

$$C_{t,j,k}^{(Sample)} = T^{-1} \textstyle\sum_{\tau=0}^{T} y_{t-j-\tau}\, y_{t-k-\tau}^T\,.$$

where integer T is the size of trailing sample. A better approach is "realized exponentially weighted moving average",

$$C_{t,j,k}^{(EWMA)} = \lambda\, y_{t-j}\, y_{t-k}^T + (1-\lambda)\, C_{(t-1),j,k}^{(EWMA)}\,,$$

where λ can be referred to as "smoothing factor" and $0 < \lambda < 1$. A best method is numeric (e.g. "particle") or analytic (e.g. "variational") Bayesian filtering,

$$C_{t,j,k}^{(BF)} = F(\, y_{t-j}\, y_{t-k}^{T} \mid C_{(t-1),j,k}^{(BF)}\, ;\, \theta_C\,)\,, \tag{A.1.3}$$

where, of Bayesian analysis of the filter, $C_{t,j,k}^{(BF)}$ is posterior, y_t is current observations as data, $C_{(t-1),j,k}^{(BF)}$ serves as prior at time t, and θ_C is a set of all parameters (may include assumptions on distributions of y_t and $C_{t,j,k}$).

A non-stationary DFM of y_t can be estimated with principal eigenpairs, i.e. pairs of principal eigenvalues and associated principal eigenvectors (or "principal components") of the VCMs $C_{t,j,j}$ of y_t, $j = 0, 1, \dots, p$. Due to random measurement errors (noises) in sample data of limited sample sizes, direct calculations of eigenpairs with individual VCMs $C_{t,j,j}$ are often not considered acceptable in actual model estimations. A better DFM estimation is to utilize the entire set of all estimated covariance matrixes $C_{t,j,k}^{(Est)}$, $j, k = 0, 1, \dots, p$.

Estimating stationary DFMs by spectral PCAs through Fourier transforms is the latest achievement in this direction (Doz & Fuleky 2020). To estimate DFM on discrete vector time-series y_t with non-stationary covariance matrixes $C_{t,j,k}$ by Eq. (A.1.1) and two time-lags j and k, we can utilize principal components (PCs) of spectral density matrixes (SDMs) obtained through conjugate two-dimensional discrete Fourier transform, or C2D-DFT.

A standard 2D-DFT on covariance matrixes $C_{t,j,k}$ is

$$S_{t,m,n}^{(O)} = \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,j,k}\; exp[-i2\pi(mj + nk)\;/\;(p+1)]\,, \tag{A.1.4}$$

where $m, n = 0, 1, \dots, p$, and superscript (O) indicates that $S_{t,m,n}^{(O)}$ results from "original" standard 2D-DFT. Inverse of the standard 2D-DFT is

$$C_{t,j,k} = (p+1)^{-2} \sum_{m=0}^{p} \sum_{n=0}^{p} S_{t,m,n}^{(O)}\; exp[+i2\pi(mj + nk)\;/\;(p+1)]\,, \tag{A.1.5}$$

where $j, k = 0, 1, \dots, p$. Since covariance matrixes $C_{t,j,k}$ by Eq. (A.1.1) are real-valued matrixes, when $n = m$, transpose of $S^{(O)}_{t,m,m}$ by Eq. (A.1.4) is

$$(S^{(O)}_{t,m,m})^T = \sum_{j=0}^{p} \sum_{k=0}^{p} C^T_{t,j,k}\ exp[-i2\pi m(j+k)\ /\ (p+1)]$$
$$= \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,k,j}\ exp[-i2\pi m(j+k)\ /\ (p+1)] = S^{(O)}_{t,m,m}\ .$$

The complex-valued matrix $S^{(O)}_{t,m,m}$, $0 \le m \le p$, equals to its non-conjugate transpose, i.e. $S^{(O)}_{t,m,m}$ is symmetric, but not Hermitian.

If, beyond the set of calculations of standard 2D-DFT above, we also calculate $S^{(O)}_{t,m,n}$ with $n = 0, (-1), \dots, (-p)$, and denote $S_{t,m,n} = S^{(O)}_{t,m,(-n)}$, or $S^{(O)}_{t,m,n} = S_{t,m,(-n)}$, the pair of standard 2D-DFT Eqs. (A.1.4) and (A.1.5) become

$$S_{t,m,n} = \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,j,k}\ exp[-i2\pi(mj - nk)\ /\ (p+1)]\ , \qquad \text{(A.1.6)}$$

where $m, n = 0, 1, \dots, p$, and

$$C_{t,j,k} = (p+1)^{-2} \sum_{m=0}^{p} \sum_{n=0}^{p} S_{t,m,n}\ exp[+i2\pi(mj - nk)\ /\ (p+1)]\ , \qquad \text{(A.1.7)}$$

where $j, k = 0, 1, \dots, p$. Let's denote A^* as conjugate of matrix A. When $n = m$, transpose of $S_{t,m,m}$ by Eq. (A.1.6) is

$$S^T_{t,m,m} = \sum_{j=0}^{p} \sum_{k=0}^{p} C^T_{t,j,k}\ exp[-i2\pi m(j-k)\ /\ (p+1)]$$
$$= \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,k,j}\ exp[-i2\pi m(j-k)\ /\ (p+1)]$$
$$= \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,j,k}\ exp[+i2\pi m(j-k)\ /\ (p+1)] = S^*_{t,m,m}\ .$$

The complex-valued matrix $S_{t,m,m}$, $0 \le m \le p$, equals to its conjugate transpose, i.e. matrix $S_{t,m,m}$ is Hermitian, while matrix $S^{(O)}_{t,m,m}$ is not.

Of Hermitian matrixes $S_{t,m,m}$, $m = 0, 1, \dots, p$, all eigenvalues are real-valued and non-negative, and complex eigenvectors associated with distinct eigenvalues are orthogonal to one another. These advantages of Hermitian matrixes make Eqs. (A.1.6) and (A.1.7) much more suitable for spectral PCA than the standard 2D-DFT. To differentiate this variant of 2D-DFT by Eqs. (A.1.6) and (A.1.7) from the original standard 2D-DFT, we name

- the two-dimensional discrete Fourier transform by Eq. (A.1.6) as "conjugate 2D-DFT", or C2D-DFT,
- the inverse by Eq. (A.1.7) as "inverse conjugate 2D-DFT", or inverse C2D-DFT, or Inv-C2D-DFT,
- matrixes $S_{t,m,n}$, $m, n = 0, 1, \dots, p$, by C2D-DFT as spectral density matrixes, or SDMs, of observed vector time-series y_t, and
- matrixes $S_{t,m,m}$ of $n = m$ as "on-diagonal" SDM while $S_{t,m,n}$ of $n \neq m$ as "off-diagonal" SDM (as if pairs $[m, n]$ are elements of a symmetric matrix).

Let's denote A^H as conjugate transpose of matrix A. Substituting Eq. (A.1.2) of symmetry into Eq. (A.1.6) of C2D-DFT yields

$$\begin{aligned} S_{t,m,n} &= \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,j,k} \; exp[-i2\pi(mj - nk)/(p+1)] \\ &= \sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,k,j}^{T} \; exp[+i2\pi(nk - mj)/(p+1)] \\ &= S_{t,n,m}^{H} \,, \end{aligned} \tag{A.1.8}$$

and when $n = m$, we have

$$S_{t,m,m} = S_{t,m,m}^{H} \,. \tag{A.1.9}$$

Conjugate transpose of SDM $S_{t,m,n}$ equals to SDM $S_{t,n,m}$, and on-diagonal SDM $S_{t,m,m}$ is Hermitian.

When applying standard (one-dimensional) discrete Fourier transform (DFT) on observed vector time-series y_t,

$$z_{t,m} = \sum_{j=0}^{p} y_{t-j} \, exp[-i2\pi mj \,/\, (p+1)] \,, \tag{A.1.10}$$

$$y_{t-j} = (p+1)^{-1} \sum_{m=0}^{p} z_{t,m} \, exp[+i2\pi mj \,/\, (p+1)] \,, \tag{A.1.11}$$

where $m, j = 0, 1, \dots, p$, Eq. (A.1.10) is DFT, and Eq. (A.1.11) is inverse DFT. Covariance matrix of complex-value stochastic vectors $z_{t,m}$ and $z_{t,n}$ in frequency domain is

$$\begin{aligned}
& W_{t,m,n} \\
&= Cov(\, z_{t,m}\,,\, z_{t,n}\,) = E(\, z_{t,m}\, z_{t,n}^{H}\,) \\
&= E(\, \textstyle\sum_{j=0}^{p} \sum_{k=0}^{p} y_{t-j}\, y_{t-k}^{T}\, exp[-i2\pi(mj - nk) \,/\, (p+1)]\,) \\
&= \textstyle\sum_{j=0}^{p} \sum_{k=0}^{p} E(\, y_{t-j}\, y_{t-k}^{T}\,)\, exp[-i2\pi(mj - nk) \,/\, (p+1)] \\
&= \textstyle\sum_{j=0}^{p} \sum_{k=0}^{p} C_{t,j,k}\; exp[-i2\pi(mj - nk) \,/\, (p+1)] \\
&= S_{t,m,n}\,,
\end{aligned} \tag{A.1.12}$$

where $m, n = 0, 1, \dots, p$. Therefore, spectral density matrixes $S_{t,m,n}$ by C2D-DFT in Eq. (A.1.6) are, actually, covariance matrixes of frequency components $z_{t,m}$ and $z_{t,n}$ of vector time-series y_t at specific frequencies.

Similar to Bayesian filtering by Eq. (A.1.3), covariance by Eq. (A.1.12) indicates that the SDMs $S_{t,m,n}$ can be estimated by a particle or variational Bayesian filtering as

$$S_{t,m,n}^{(BF)} = F(\, z_{t,m}\, z_{t,n}^{H} \,|\, S_{(t-1),m,n}^{(BF)}\,;\ \theta_S\,)\,, \tag{A.1.13}$$

where, of Bayesian analysis of the filter, $S_{t,m,n}^{(BF)}$ is posterior, $z_{t,m}$ is current information as data, $S_{(t-1),m,,n}^{(BF)}$ serves as prior at time t, and θ_S is a set of all parameters (may include assumptions on distributions of $z_{t,m}$ and $S_{t,m,n}$).

A.2. Eigenpairs of VCMs and SDMs

According to long established academic research, covariance matrixes (VCMs and ACMs) defined by Eq. (A.1.1), $C_{t,j,k} = C_{t,k,j}^{T}$, $0 \le j, k \le p$, can be decomposed into product of 3 square matrixes as,

$$C_{t,j,k} = G_{t,j}^{(C)}\, Q_{t,j,k}^{(C)}\, (G_{t,k}^{(C)})^{T}\,, \tag{A.2.1}$$

where superscript (C) indicates that the component matrixes, $G_{t,j}^{(C)}$, $Q_{t,j,k}^{(C)}$ and $G_{t,k}^{(C)}$, are of covariance matrix $C_{t,j,k}$.

It is classic and well-known that, when $k = j$, the symmetric VCM $C_{t,j,j}$ can be decomposed into

$$C_{t,j,j} = G_{t,j}^{(C)}\, Q_{t,j}^{(C)}\, (G_{t,j}^{(C)})^T = G_{t,j}^{(C)}\, Q_{t,j,j}^{(C)}\, (G_{t,j}^{(C)})^T , \qquad \text{(A.2.2)}$$

where $j = 0, 1, \cdots, p$, the matrix at middle, $Q_{t,j}^{(C)} = Q_{t,j,j}^{(C)}$, is diagonal matrix with elements on the diagonal are entire set of non-negative eigenvalues of $C_{t,j,j}$, and columns of the square matrix at left, $G_{t,j}^{(C)}$, are entire set of correspondent eigenvectors of $C_{t,j,j}$.

Of symmetric matrixes, eigenvectors associated with distinct eigenvalues are orthogonal to one another. If eigenvalues are not distinct, eigenvectors associated with repeated eigenvalues can be transformed into orthogonal to one another and to all others. Furthermore, all eigenvectors can be rescaled into orthonormal ones as

$$(G_{t,j}^{(C)})^T\, G_{t,j}^{(C)} = I , \qquad \text{(A.2.3)}$$

where I is identity matrix and thus inverse of $G_{t,j}^{(C)}$ is

$$(G_{t,j}^{(C)})^{-1} = (G_{t,j}^{(C)})^T , \qquad \text{(A.2.4)}$$

i.e. inverse of orthonormal $G_{t,j}^{(C)}$ is transpose of $G_{t,j}^{(C)}$. Therefore, in this paper, we assume all eigenvectors are orthonormal.

Applying orthonormal eigenvector matrixes $G_{t,j}^{(C)}$ and $G_{t,k}^{(C)}$ to Eq. (A.2.1), square coefficient matrix $Q_{t,j,k}^{(C)}$ can be calculated by covariance matrix $C_{t,j,k}$ and eigenvectors $G_{t,j}^{(C)}$ and $G_{t,k}^{(C)}$ as

$$Q_{t,j,k}^{(C)} = (G_{t,j}^{(C)})^T\, C_{t,j,k}\, G_{t,k}^{(C)} . \qquad \text{(A.2.5)}$$

Such “quadratic projection” Eq. (A.2.5) is useful in our PCA estimations.

We can define an unobserved component vector s_{t-j} by observed data vector y_{t-j} and calculated eigenvector matrix $G_{t,j}^{(C)}$ as

$$s_{t-j} = (G_{t,j}^{(C)})^T y_{t-j} , \tag{A.2.6}$$

where each individual element of s_{t-j} , can be called an “eigen score”, is a linear combination of all elements of y_{t-j} . And each individual element of y_{t-j} is a linear combination of all eigen score elements of s_{t-j} as

$$y_{t-j} = G_{t,j}^{(C)} s_{t-j} . \tag{A.2.7}$$

Substituting Eq. (A.2.7) into Eq. (A.1.1) for $k = j$ and comparing with Eq. (A.2.2) yields,

$$C_{t,j,j} = E(y_{t-j} y_{t-j}^T) = G_{t,j}^{(C)} E(s_{t-j} s_{t-j}^T) (G_{t,j}^{(C)})^T = G_{t,j}^{(C)} Q_{t,j}^{(C)} (G_{t,j}^{(C)})^T , \tag{A.2.8}$$

i.e. diagonal elements of $Q_{t,j}^{(C)} = Q_{t,j,j}^{(C)}$, or eigenvalues, are variances of unobserved eigen scores s_{t-j} . Substituting Eq. (A.2.7) into Eq. (A.1.1) for $k \neq j$ and comparing with Eq. (A.2.1) yields

$$C_{t,j,k} = E(y_{t-j} y_{t-k}^T) = G_{t,j}^{(C)} E(s_{t-j} s_{t-k}^T) (G_{t,k}^{(C)})^T = G_{t,j}^{(C)} Q_{t,j,k}^{(C)} (G_{t,k}^{(C)})^T , \tag{A.2.9}$$

i.e. square matrix $Q_{t,j,k}^{(C)}$ is vector autocovariance (VAR) matrix of unobserved eigen scores s_{t-j} and s_{t-k} .

As discussed in Appendix A.1, spectral density matrixes $S_{t,m,n}$ obtained by applying conjugate 2D-DFT on covariance matrixes $C_{t,j,k}$ of observed vector time-series y_t with Eq. (A.1.6) are, actually, covariance matrixes of frequency components $z_{t,m}$ and $z_{t,n}$ of y_t obtained by applying DFT on vector time-series y_t with Eq. (A.1.10).

Then, on-diagonal SDMs $S_{t,m,m}$, $m = 0, 1, \dots, p$, are complex-valued Hermitian equivalents of real-valued symmetric VCMs $C_{t,j,j}$, have non-negative real eigenvalues and correspondent complex eigenvectors, and can be decomposed by eigenpairs in a way same as Eq. (A.2.2) as

$$S_{t,m,m} = G_{t,m}^{(S)} Q_{t,m}^{(S)} (G_{t,m}^{(S)})^H = G_{t,m}^{(S)} Q_{t,m,m}^{(S)} (G_{t,m}^{(S)})^H \, , \qquad \text{(A.2.10)}$$

where superscript (S) indicates that the square component matrixes $Q_{t,m}^{(S)}$ and $G_{t,m}^{(S)}$ are of spectral density matrix $S_{t,m,m}$. The matrix at middle, $Q_{t,m}^{(S)} = Q_{t,m,m}^{(S)}$, is diagonal matrix with elements on the diagonal are entire set of non-negative real eigenvalues of $S_{t,m,m}$, and columns of the square matrix at left, $G_{t,m}^{(S)}$, are entire set of correspondent complex eigenvectors of $S_{t,m,m}$. Similar to Eqs. (A.2.3) and (A.2.4), we assume all complex eigenvectors are orthonormal,

$$(G_{t,m}^{(S)})^H G_{t,m}^{(S)} = I \, , \qquad \text{(A.2.11)}$$

and inverse of $G_{t,m}^{(S)}$ is

$$(G_{t,m}^{(S)})^{-1} = (G_{t,m}^{(S)})^H \, , \qquad \text{(A.2.12)}$$

i.e. inverse of complex orthonormal $G_{t,m}^{(S)}$ is conjugate transpose of $G_{t,m}^{(S)}$.

On the other hand, off-diagonal SDMs $S_{t,m,n}$, $n \neq m$, are complex-valued equivalents of real-valued ACMs $C_{t,j,k}$, can be decomposed by eigenpairs in a way same as Eq. (A.2.1) as

$$S_{t,m,n} = G_{t,m}^{(S)} Q_{t,m,n}^{(S)} (G_{t,n}^{(S)})^H \, . \qquad \text{(A.2.13)}$$

where $m, n = 0, 1, \dots, p$. Similar to Eq. (A.2.5), square coefficient matrix $Q_{t,m,n}^{(S)}$ can be calculated by SDM $S_{t,m,n}$ and orthonormal eigenvector matrixes $G_{t,m}^{(S)}$ and $G_{t,n}^{(S)}$ as

$$Q_{t,m,n}^{(S)} = (G_{t,m}^{(S)})^H S_{t,m,n} G_{t,n}^{(S)} \, . \qquad \text{(A.2.14)}$$

This quadratic projection Eq. (A.2.14) is useful in our spectral PCA estimations.

A.3. Spectral PCA

Three equations Eqs. (A.2.10), (A.2.14) and (A.2.13), each composed of entire set of spectral eigenpairs in frequency domain, are starting point of this section and, therefore, repeated here for ease of discussion:

$$S_{t,m,m} = G_{t,m}^{(S)}\, Q_{t,m}^{(S)}\, (G_{t,m}^{(S)})^{H} = G_{t,m}^{(S)}\, Q_{t,m,m}^{(S)}\, (G_{t,m}^{(S)})^{H}\ , \tag{A.3.1}$$

$$Q_{t,m,n}^{(S)} = (G_{t,m}^{(S)})^{H}\, S_{t,m,n}\, G_{t,n}^{(S)}\ , \tag{A.3.2}$$

$$S_{t,m,n} = G_{t,m}^{(S)}\, Q_{t,m,n}^{(S)}\, (G_{t,n}^{(S)})^{H}\ , \tag{A.3.3}$$

where $S_{t,m,n}$ is spectral density matrix (SDM), $Q_{t,m}^{(S)}$ is diagonal matrix with all eigenvalues of $S_{t,m,m}$ (when $n = m$) as its diagonal, columns of matrix $G_{t,m}^{(S)}$ are all correspondent orthonormal eigenvectors of $S_{t,m,m}$, $Q_{t,m,n}^{(S)}$ is coefficient matrix of $S_{t,m,n}$, and $m, n = 0, 1, \dots, p$.

One way to make spectral principal components analysis, or spectral PCA, on SDMs $S_{t,m,n}$ is to replace the entire set of eigenpairs in the three equations above by a (much) smaller set of "principal eigenpairs" associated with the largest eigenvalues. All other eigenpairs associated with smaller eigenvalues may well be irrelevant to unobserved common factors underlying the vector time-series y_t .

To carry out spectral PCA on those on-diagonal SDMs $S_{t,m,m}$ in Eq. (A.3.1), assume that

- all non-negative real eigenvalues of Hermitian $S_{t,m,m}$ are orderly placed from largest to smallest on the diagonal of matrix $Q_{t,m}^{(S)}$,
- all complex orthonormal eigenvectors of $S_{t,m,m}$ (associated with the sorted

eigenvalues) are orderly placed as columns in square matrix $G_{t,m}^{(S)}$ from left to right, and

- integer L_P as number of largest eigenvalues to be included in spectral PCA.

Then $L_P \times L_P$ upper-left sub-matrix $P_{t,m}^{(S)}$ of $Q_{t,m}^{(S)}$ contains L_P largest eigenvalues or "principal eigenvalues" of $S_{t,m,m}$, and L_P-column left sub-matrix $H_{t,m}^{(S)}$ of $G_{t,m}^{(S)}$ contains L_P "principal eigenvectors" of $S_{t,m,m}$ associated with the principal eigenvalues. When replacing matrixes $Q_{t,m}^{(S)}$ and $G_{t,m}^{(S)}$ of all eigenpairs by their sub-matrixes $P_{t,m}^{(S)}$ and $H_{t,m}^{(S)}$ of principal eigenpairs, Eq. (A.3.1) becomes a three-matrix product based only on the L_P principal eigenpairs as

$$S_{t,m,m}^{(PC)} = H_{t,m}^{(S)} P_{t,m}^{(S)} (H_{t,m}^{(S)})^H , \tag{A.3.4}$$

where superscript (PC) indicates Hermitian matrix $S_{t,m,m}^{(PC)}$ is a "principal components (PC) representation" of on-diagonal SDM $S_{t,m,m}$, $m = 0, 1, \dots, p$.

The matrix of L_P orthonormal principal eigenvector columns, $H_{t,m}^{(S)}$, is no longer a square matrix, with its number of columns L_P (much) smaller than its number of rows, and

$$(H_{t,m}^{(S)})^H H_{t,m}^{(S)} = I , \tag{A.3.5}$$

where I is $L_P \times L_P$ identity matrix.

For those off-diagonal SDMs $S_{t,m,n}$, $n \neq m$, when replacing matrixes $G_{t,m}^{(S)}$ and $G_{t,n}^{(S)}$ of all eigenvectors by their sub-matrixes $H_{t,m}^{(S)}$ and $H_{t,n}^{(S)}$ of principal eigenvectors, Eq. (A.3.2) becomes a quadratic projection based only on the principal eigenpairs as

$$P_{t,m,n}^{(S)} = (H_{t,m}^{(S)})^H S_{t,m,n} H_{t,n}^{(S)} , \tag{A.3.6}$$

where the resulted $L_P \times L_P$ matrix $P_{t,m,n}^{(S)}$ is, actually, a sub-matrix of the coefficient matrix

$Q^{(S)}_{t,m,n}$ in Eqs. (A.3.2) and (A.3.3). The conjugate transpose of $P^{(S)}_{t,m,n}$,

$$(P^{(S)}_{t,m,n})^H = (H^{(S)}_{t,n})^H\, S^H_{t,m,n}\, H^{(S)}_{t,m} = (H^{(S)}_{t,n})^H\, S_{t,n,m}\, H^{(S)}_{t,m} = P^{(S)}_{t,n,m}\ , \qquad \text{(A.3.7)}$$

is coefficient matrix of conjugate transpose of $S_{t,m,n}$.

After replacing all three matrixes $G^{(S)}_{t,m}$, $Q^{(S)}_{t,m,n}$ and $G^{(S)}_{t,n}$ of all eigenpairs by their sub-matrixes $H^{(S)}_{t,m}$, $P^{(S)}_{t,m}$ and $H^{(S)}_{t,n}$ of principal eigenpairs, Eq. (A.3.3) becomes

$$S^{(PC)}_{t,m,n} = H^{(S)}_{t,m}\, P^{(S)}_{t,m,n}\, (H^{(S)}_{t,n})^H\ , \qquad \text{(A.3.8)}$$

where complex-valued square matrix $S^{(PC)}_{t,m,n}$ is principal components (PC) representation of SDM $S_{t,m,n}$, $0 \le m,n \le p$. The conjugate transpose of $S^{(PC)}_{t,m,n}$,

$$(S^{(PC)}_{t,m,n})^H = H^{(S)}_{t,n}\, (P^{(S)}_{t,m,n})^H\, (H^{(S)}_{t,m})^H = H^{(S)}_{t,n}\, P^{(S)}_{t,n,m}\, (H^{(S)}_{t,m})^H = S^{(PC)}_{t,n,m}\ , \qquad \text{(A.3.9)}$$

is PC representation of conjugate transpose of SDM $S_{t,m,n}$.

As discussed in Appendix A.1, spectral density matrixes $S_{t,m,n}$ obtained by applying C2D-DFT on covariance matrixes $C_{t,j,k}$ of observed vector time-series y_t are, actually, covariance matrixes of frequency components $z_{t,m}$ and $z_{t,n}$ of y_t obtained by applying DFT on vector time-series y_t as

$$z_{t,m} = \sum_{j=0}^{p} y_{t-j}\, exp[-i2\pi mj\ /\ (p+1)]\ , \qquad \text{(A.3.10)}$$

where $m = 0, 1, \dots, p$.

Based on eigen score equations Eq. (A.2.6), “spectral principal eigen score vector” $f^{(S)}_{t,m}$ can be defined by frequency components $z_{t,m}$ in Eq. (A.3.10) and spectral principal eigenvector matrix $H^{(S)}_{t,m}$ in Eq. (A.3.4) as

$$f^{(S)}_{t,m} = (H^{(S)}_{t,m})^H\, z_{t,m}\ , \qquad \text{(A.3.11)}$$

where each element of principal eigen score vector $f_{t,m}^{(S)}$ is a linear combination of all elements of $z_{t,m}$. According to Eqs. (A.2.8) and (A.2.9), variances of principal eigen scores $f_{t,m}^{(S)}$ are principal eigenvalues on diagonal of matrix $P_{t,m}^{(S)}$ in Eq. (A.3.4), and covariance of $f_{t,m}^{(S)}$ and $f_{t,n}^{(S)}$, $n \neq m$, is coefficient matrix $P_{t,m,n}^{(S)}$ by Eq. (A.3.6).

Since number of columns L_P of principal eigenvector matrix $H_{t,m}^{(S)}$ is (much) smaller than its number of rows, frequency components $z_{t,m}$ cannot be fully recovered by principal eigen scores $f_{t,m}^{(S)}$ in a way as Eq. (A.2.7). Instead,

$$z_{t,m}^{(PC)} = H_{t,m}^{(S)} f_{t,m}^{(S)} = H_{t,m}^{(S)} (H_{t,m}^{(S)})^H z_{t,m} , \tag{A.3.12}$$

where $z_{t,m}^{(PC)}$ is principal components (PC) representation of frequency components $z_{t,m}$, and low (L_P) rank matrix $H_{t,m}^{(S)} (H_{t,m}^{(S)})^H$ projects frequency components $z_{t,m}$ into its PC representation $z_{t,m}^{(PC)}$.

In actual spectral PCA process, one computes principal eigenvalues and principal eigenvectors directly, without any need of getting all other eigenvalues and eigenvectors at all.

A.4. PCA after Inv-C2D-DFT

Four equations Eqs. (A.3.4), (A.3.8), (A.3.11) and (A.3.12), each based only on selected spectral principal eigenpairs in frequency domain, are starting point of this section and, therefore, repeated here for ease of discussion:

$$S_{t,m,m}^{(PC)} = H_{t,m}^{(S)} P_{t,m}^{(S)} (H_{t,m}^{(S)})^H , \tag{A.4.1}$$

$$S_{t,m,n}^{(PC)} = H_{t,m}^{(S)} P_{t,m,n}^{(S)} (H_{t,n}^{(S)})^H , \tag{A.4.2}$$

$$f_{t,m}^{(S)} = (H_{t,m}^{(S)})^H z_{t,m} , \tag{A.4.3}$$

$$z_{t,m}^{(PC)} = H_{t,m}^{(S)} (H_{t,m}^{(S)})^H z_{t,m} , \tag{A.4.4}$$

where $P_{t,m}^{(S)}$ is diagonal matrix composed of principal eigenvalues of on-diagonal SDM $S_{t,m,m}$, columns of $H_{t,m}^{(S)}$ are principal eigenvectors of $S_{t,m,m}$, $f_{t,m}^{(S)}$ is vector of principal eigen scores, $S_{t,m,n}^{(PC)}$ is principal components (PC) representation of SDM $S_{t,m,n}$, and $z_{t,m}^{(PC)}$ is PC representation of frequency components $z_{t,m}$ of vector time-series y_t. These matrixes and vectors composed of, or based on, spectral principal eigenpairs in frequency domain are not directly utilized by non-stationary DFM estimation in time domain.

Fortunately, dimension-reduced representations of covariance matrixes $C_{t,j,k}$ in Eq. (A.1.1) can be obtained by applying inverse conjugate 2D-DFT or Inv-C2D-DFT, in Eq. (A.1.7), on PC representations $S_{t,m,n}^{(PC)}$ of SDMs in Eq. (A.4.2) as

$$\begin{aligned} &C_{t,j,k}^{(SPC)} \\ &= (p+1)^{-2} \textstyle\sum_{m=0}^{p} \sum_{n=0}^{p} S_{t,m,n}^{(PC)} \, exp[+i2\pi(mj - nk)/(p+1)] \end{aligned} \tag{A.4.5}$$

where superscript (SPC) indicates that (complex-valued) $C_{t,j,k}^{(SPC)}$ are "spectral PC (or SPC) representations" of real-valued covariance matrixes $C_{t,j,k}$, $j, k = 0, 1, \dots, p$. Conjugate transposes of $C_{t,j,k}^{(SPC)}$,

$$\begin{aligned} &(C_{t,j,k}^{(SPC)})^H \\ &= (p+1)^{-2} \textstyle\sum_{m=0}^{p} \sum_{n=0}^{p} (S_{t,m,n}^{(P)})^H \, exp[-i2\pi(mj - nk)/(p+1)] \\ &= (p+1)^{-2} \textstyle\sum_{m=0}^{p} \sum_{n=0}^{p} S_{t,n,m}^{(P)} \, exp[+i2\pi(nk - mj)/(p+1)] \\ &= C_{t,k,j}^{(SPC)} , \end{aligned} \tag{A.4.6}$$

is SPC representation of transpose of covariance matrix.

SPC representation $C_{t,j,j}^{(SPC)}$ of variance-covariance matrix (VCM) $C_{t,j,j}$ (when $k = j$)

by Eq. (A.4.5) is linear combination of all complex-valued PC representations, $S_{t,m,n}^{(PC)}$, of SDMs. As a result,

- $C_{t,j,j}^{(SPC)}$ could be a complex Hermitian matrix (instead of a real symmetric matrix as $C_{t,j,j}$ itself),
- number of positive eigenvalues of $C_{t,j,j}^{(SPC)}$ could be larger than L_P (instead of L_P as every individual $S_{t,m,m}^{(PC)}$),
- Inv-C2D-DFT by Eq. (A.4.5) does not give rise to actual sets of principal eigenpairs of $C_{t,j,j}^{(SPC)}$,

where $j = 0, 1, \dots, p$. Therefore, one can take only the real part of $C_{t,j,k}^{(SPC)}$,

$$C_{t,j,k}^{(Re|SPC)} = Re(\, C_{t,j,k}^{(SPC)}\,)\,, \tag{A.4.7}$$

as "applicable SPC representations" of covariance matrixes $C_{t,j,k}$. Transpose of $C_{t,j,k}^{(Re|SPC)}$,

$$(C_{t,j,k}^{(Re|SPC)})^T = C_{t,k,j}^{(Re|SPC)}\,, \tag{A.4.8}$$

is real-valued counterpart of complex-valued relationship in Eq. (A.4.6).

For $k = j$, applicable SPC representations $C_{t,j,j}^{(Re|SPC)}$ of VCMs $C_{t,j,j}$ are real-valued symmetric matrixes. One can apply additional PCA in time-domain on $C_{t,j,j}^{(Re|SPC)}$ to get L_P principal eigenpairs of $C_{t,j,j}^{(Re|SPC)}$ representing L_P principal eigenpairs of covariance matrix $C_{t,j,j}$,

$$C_{t,j,j}^{(PC)} = H_{t,j}^{(C)}\, P_{t,j}^{(C)}\, (H_{t,j}^{(C)})^T = H_{t,j}^{(C)}\, P_{t,j,j}^{(C)}\, (H_{t,j}^{(C)})^T\,, \tag{A.4.9}$$

where superscript (PC) indicates that $C_{t,j,j}^{(PC)}$ is "principal components (PC) representation" of $C_{t,j,j}$, the matrix at middle, $P_{t,j}^{(C)} = P_{t,j,j}^{(C)}$, is diagonal matrix with elements on the diagonal are principal eigenvalues of $C_{t,j,j}^{(Re|SPC)}$, columns of the matrix at left, $H_{t,j}^{(C)}$, are correspondent principal eigenvectors of $C_{t,j,j}^{(Re|SPC)}$, and all principal eigenvectors are orthonormal,

$$(H_{t,j}^{(C)})^T H_{t,j}^{(C)} = I , \qquad \text{(A.4.10)}$$

where $j = 0, 1, \cdots, p$.

For $k \neq j$, calculate "principal coefficient matrix" $P_{t,j,k}^{(C)}$ by quadratic projection similar to Eq. (A.2.5) as

$$P_{t,j,k}^{(C)} = (H_{t,j}^{(C)})^T C_{t,j,k}^{(Re|SPC)} H_{t,k}^{(C)} , \qquad \text{(A.4.11)}$$

and, then, PC representation of vector autocovariance matrix $C_{t,j,k}$ is

$$C_{t,j,k}^{(PC)} = H_{t,j}^{(C)} P_{t,j,k}^{(C)} (H_{t,k}^{(C)})^T , \qquad \text{(A.4.12)}$$

where $j, k = 0, 1, \cdots, p$.

Applying inverse DFT (Inv-DFT) shown in Eq. (A.1.11) on PC representations $z_{t,m}^{(PC)}$ of frequency components $z_{t,m}$ of observed vector time-series y_t obtained by Eq. (A.4.4) yields

$$y_{t-j}^{(SPC)} = (p+1)^{-1} \sum_{m=0}^{p} z_{t,m}^{(PC)} exp[+i2\pi mj \,/\, (p+1)] , \qquad \text{(A.4.13)}$$

where $y_{t-j}^{(SPC)}$, is SPC representation of data vectors y_{t-j}, $j = 0, 1, \cdots, p$.

With orthonormal principal eigenvector matrix $H_{t,j}^{(C)}$ of covariance matrix $C_{t,j,j}$ in Eq. (A.4.11) and SPC representation $y_{t-j}^{(SPC)}$ of vector time-series y_{t-j} in Eq. (A.4.13), principal eigen scores $f_{t-j}^{(C)}$ of y_{t-j} can be calculated as a time domain equivalent of Eq. (A.4.3) as

$$f_{t-j}^{(C)} = (H_{t,j}^{(C)})^T y_{t-j}^{(SPC)} , \qquad \text{(A.4.14)}$$

where $j = 0, 1, \cdots, p$. Principal eigenvalues $P_{t,j}^{(C)}$ in Eq. (A.4.9) are variances of principal eigen scores $f_{t-j}^{(C)}$, and coefficient matrixes $P_{t,j,k}^{(C)}$ by Eq. (A.4.11) are vector autocovariances of principal eigen scores $f_{t-j}^{(C)}$ and $f_{t-k}^{(C)}$.

In non-stationary DFM estimation process with steps discussed in Appendixes A.1 through A.4, principal eigenpairs in matrixes $P_{t,j}^{(C)}$, $H_{t,j}^{(C)}$ and $P_{t,j,k}^{(C)}$ of individual covariance matrix $C_{t,j,k}$ shown in Eqs. (A.4.9) and (A.4.11) are not directly calculated using sample estimate of single covariance matrix $C_{t,j,k}$. Every set of principal eigenpairs are estimated on combination and re-combination of all covariance matrixes, through conjugate 2D-DFT, spectral PCA, inverse C2D-DFT, additional PCA in time domain, and classic DFT and inverse DFT.

Finally, in a non-stationary dynamic factor model estimation, based on the approaches discussed by Forni, Hallin, Lippi and Reichlin (2000),

- dynamic factor loadings matrixes $X_{t,j}^{(p_1)}$,
- common factor scores f_t,
- dynamic factor VAR matrixes $A_{t,k}^{(p_2)}$,
- variances V_{t-j} of factor scores,
- vector autocovariances $V_{(t-j),(t-k)}$ of factor scores,
- factor filter coefficient matrixes $S_{t,j}$, and
- DFM-based VAR matrixes $M^{(L)}$ and $M_{t,k}^{(R)}$

are estimated by results of spectral PCA with conjugate 2D-DFT:

- principal eigenvectors $H_{t,j}^{(C)}$ in Eq. (A.4.9),
- principal eigen scores $f_{t-j}^{(C)}$ by Eq. (A.4.14),
- principal eigenvalues $P_{t,j}^{(C)} = P_{t,j,j}^{(C)}$ in Eq. (A.4.9) and
- (principal) coefficient matrixes $P_{t,j,k}^{(C)}$ by Eq. (A.4.11),

where additional PCA in time domain and classic DFT and inverse DFT are also involved.

A final footnote is as follows. The matrix relations involving transpose, symmetry, conjugate transpose and Hermitian matrix can be utilized to avoid many repeated calculations in the process of estimating non-stationary DFM by spectral PCA with C2D-DFT.

References

M. Barigozzi (2024). "Dynamic Factor Models". Università di Bologna, October 2024. https://www.barigozzi.eu/MB_DFM_complete_slides.pdf.

C. Doz and P. Fuleky (2020). "Chapter 2, Dynamic Factor Models" In *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, Ed. P. Fuleky, Advanced Studies in Theoretical and Applied Econometrics, Volume 52. Springer.

i4cast LLC (2023). "Introduction to Multi-step Forecast of Multivariate Volatility with Dynamic Factor Model". https://github.com/i4cast/aws/blob/main/dynamic_factor_variance-covariance_model/publication/multi-step_forecast_of_multivariate_volatility_with_dynamic_factor_model.pdf, an introduction to DFVCM algorithm on AWS: https://aws.amazon.com/marketplace/pp/prodview-yvaulquatt3v2?sr=0.

M. Forni, M. Hallin, M. Lippi and L. Reichlin (2000). "The Generalized Dynamic Factor Model: Identification and Estimation". The Review of Economics and Statistics, November 2000, 82 (4), pp. 540–554.

[EOF]