Chapter 4

Dynamic and spectral factor models

Recall the **linear model** $\mathsf{y} = Wx + \mathsf{z}$ of (3.7). The intention there was to linearly relate the set of r-variate independent samples $\{y_l\}$, $l = 1, \ldots, n$, which are thought to be **realizations** of a **vector random variable** y to the corresponding set of q-variate samples $\{x_l\}$, $l = 1, \ldots, n$. However, both $\{x_l\}$ and $\{y_l\}$ were measured and available in a given dataset.

Now, contrast the linear model with the **factor model** $\mathsf{y} = W\mathsf{x} + \mathsf{z}$ of (3.11), where x is a q-variate hidden or **latent vector random variable** while y is an r-variate **measured vector random variable**.

The noticeable similarities between the linear model and the factor model are assumptions that $r > q$, transformation matrix W is non-random, and z is an r-variate vector random variable with uncorrelated components.

In this chapter, the assumption of Chapter 2 that a sequence of vector random variables $\{\mathsf{y}_t\}$ are temporally correlated is underpinned. Thence, based on the motivations presented in Chapter 1, existence of a q-variate time series $\{\mathsf{x}_t\}$ which gets transformed by a **non-stochastic** matrix $\{W_t\}$ to obtain an r-variate time series $\{\mathsf{y}_t\}$ $\forall t \in \mathbb{Z}$ is assumed. The objective of this chapter is to define such a model and enable it for learning problems.

In Section 4.1, the time-domain definition of the **dynamic factor model** and the **commonalities** it represents are defined. In doing so, the assumptions made with respect to the model are emphasized and the relations between the parameters of the model, viz., the ***acvf***s of the **measured**, **latent**, and **idiosyncratic time series** are analyzed. Then, the dynamic factor model is defined.

In Section 4.2, the analysis is switched to **Fourier-domain**: Frequency-domain counterparts of the measured, latent, and idiosyncratic time series are defined and the frequency-domain equivalent of the dynamic factor model called the **spectral factor model** is defined.

The following situates the developments in this chapter with respect to the state-of-the-art:

▷ Definition 4.2 and Definition 4.3 define the dynamic and spectral factor models, respectively.
These definitions include all model assumptions and the model objectives. In [100, 104, 43, 36], both time and frequency domain analyses are called "dynamic

factor model"; for convenience, "spectral factor model" is a term introduced here to emphasize the frequency-domain analysis. The properties and assumptions of interacting linear processes of the model used here are standard practice in literature.

▷ Definition 4.1 introduces commonalities; relations (4.9) and (4.18) state the criterion for inheriting them from the measured variables.
Unlike in the existing literature, cross-correlations are emphasized in the definitions here of dynamic and spectral factor models through the concept of commonalities. Also, the existing literature does not specifically relate the dynamic transformation to the commonalities nor its maximal inheritance as defined here.

4.1 Dynamic factor model

A **multivariate time series model** is to be designed where a q-variate **latent time series** $\{\mathsf{x}_t\}$ is transformed by a sequence of $r \times q$ non-stochastic **transformation matrices** $\{W_t\}$ to a measured r-variate time series $\{\mathsf{y}_t\}$, where $r > q$. Only those actions of W_t in transforming x_t to y_t that is intuitively appealing, theoretically valid, practically feasible, and analytically sound will be allowed. In the simplest of forms, such a time series model may be written as $\mathsf{y}_t = f(W_t, \mathsf{x}_t) + \mathsf{z}_t$, where f is some linear function of $\{W_t\}$ and $\{\mathsf{x}_t\}$ $\forall t \in \mathbb{Z}$, where $\{\mathsf{z}_t\}$ is a vector random variable independent of $\{\mathsf{x}_t\}$ that offers itself as the error in the transformation.

It is also to be ensured that the transformation will take advantage of the **frequency-based** techniques discussed in Chapter 2. In that case, existence of the **spectral density function** of $f(W_t, \mathsf{x}_t)$ is a necessity. As discussed in Section 2.4, for a weakly stationary vector random variable sequence $\{\mathsf{x}_t\}$, the Fourier transform of any linear relation $f(W_t, \mathsf{x}_t)$ between W_t and x_t does not exist for no guarantee of $\sum_{t\in\mathbb{Z}} |f(W_t, \mathsf{x}_t)| < \infty$ could be made. But as long as the *acvf* of $f(W_t, \mathsf{x}_t)$ exists and that *acvf* is absolutely summable as in (2.20), techniques of Fourier transform could be pushed.

Take a look at one of the simplest linear operations for $f(W_t, \mathsf{x}_t) \triangleq \mathsf{v}_t$, which is an r-variate vector random obtained when W_t is convolved with x_t, i.e., $\forall t \in \mathbb{Z}$,

$$\mathsf{v}_t = \sum_{j\in\mathbb{Z}} W_j \mathsf{x}_{t-j}. \tag{4.1}$$

If both $\{W_t\}$ and the *acvf* $\{\Gamma_h^{\mathsf{x}}\} \forall h \in \mathbb{Z}$ of x_t are absolutely summable, then v_t according to (4.1) exists; refer Theorem 2.7.1 of [18] for this result. Let r-variate linear processes $\{\mathsf{y}_t\}$, $\{\mathsf{v}_t\}$, and $\{\mathsf{z}_t\}$ be related according to

$$\mathsf{y}_t = \mathsf{v}_t + \mathsf{z}_t. \tag{4.2}$$

Further, if v_t and z_t are independent, then they have their *acvf*s related as

$$\Gamma_h^{\mathsf{y}} = \Gamma_h^{\mathsf{v}} + \Gamma_h^{\mathsf{z}}, \tag{4.3}$$

where $\forall h \in \mathbb{Z}$ is the lag parameter of the *acvf*s. It is further assumed that

$$\mathrm{rank}(\Gamma_h^{\mathsf{z}}) = r. \tag{4.4}$$

Thus, the measured r-variate vector random variable y_t is assumed to be obtained by adding two independent r-variate vector random variables v_t and z_t. And, v_t is a **dynamic transformation** of a latent q-variate vector random variable x_t as per (4.1).

Recall that Chapter 1 hoped to dynamically transform a latent vector random variable of known or presumed characteristics and that the dynamic transformation is the one that is unknown. The similarity between the form of dynamic transformation in (4.1) and the form of **linear process** defined in Definition 2.11 is evident. This similarity entices to assume that $\{\mathsf{x}_t\}$ is a q-variate zero mean white noise and that

$$\sum_t |W_t| < \infty \quad \forall t \in \mathbb{Z}. \tag{4.5}$$

It will deliver a $\{\mathsf{v}_t\}$ that is a linear process resulting from a linear transformation of $\{\mathsf{x}_t\}$ by the sequence of parameters $\{W_t\}$. Further, to simplify the analysis, it is assumed that $\{\mathsf{x}_t\}$ is a unit variance white noise process, i.e.,

$$\Gamma_h^{\mathsf{x}} = I_q \quad \forall h \in \mathbb{Z}. \tag{4.6}$$

Such an assumption is admissible because it is not intended to estimate Γ_h^{x} anyway. Then, referring back again to Definition 2.11, it is easy to see that

$$\Gamma_h^{\mathsf{v}} = \sum_{j \in \mathbb{Z}} W_{j+h} W_j' \quad \forall h \in \mathbb{Z}, \tag{4.7}$$

and

$$\text{rank}(\Gamma_h^{\mathsf{v}}) = q. \tag{4.8}$$

The objective is to enable $\{\mathsf{v}_t\}$ to maximally inherit the commonalities in the measured time series $\{\mathsf{y}_t\}$. And, in Chapter 1, commonalities of the measured time series $\{\mathsf{y}_t\}$ were regarded to be the temporal covariations of the r measured components of $\mathsf{y}_t = [\mathsf{y}_{1_t}\, \mathsf{y}_{2_t}\, \ldots\, \mathsf{y}_{r_t}]'$. A good measure of the commonalities should be the expected value of a suitable function combining the r random variables, e.g., their mean product. Now, the following definition is arrived at:

Definition 4.1. *For a weakly stationary time series $\{\mathsf{y}_t\}$, the **commonalities** are the off-diagonal elements of its acvf Γ_h^{y}.*

Appropriateness of Definition 4.1: Cross-covariances describe all the mutual characteristics of the components of a zero-mean multivariate time series linear process. The pairwise commonality between any two components y_{i_t} and y_{j_t} are the off-diagonal elements of the *acvf* Γ_h^{y} of the measured time series $\{\mathsf{y}_t\}$.

In Chapter 1, it was envisaged to estimate parameters of a model that will maximize the measured commonalities. Earlier in this section, the role of measured cross-covariances as a suitable measure of the commonalities was confirmed. As a result, the commonalities are retained in the cross-covariance terms of Γ_h^{v} upon a dynamic transformation of $\{\mathsf{y}_t\}$ to $\{\mathsf{v}_t\}$ as discussed earlier in this section using $\{W_t\}$. the proposed measure for the inheritance of the commonalities of Γ^{y}_h using Γ^{v}_h is the sum of square differences of the covariances across all measured dimensions and lags, i.e.,

$$g = \sum_{h \in \mathbb{Z}} \|\Gamma^{\mathsf{v}}_h - \Gamma^{\mathsf{y}}_h\|^2_F. \tag{4.9}$$

Appropriateness of g: These reason the choice of the quality of approximation of the commonalities:

1. Since $\Gamma^{\mathsf{y}}_h - \Gamma^{\mathsf{v}}_h$ is positive definite and Γ^{v}_h is of lower rank than positive definite Γ^{y}_h, trace of Γ^{v}_h will also be lower than Γ^{y}_h, i.e., the unique variance terms of Γ^{y}_h will be also affected and has to approximated in a low-rank sense.
2. It has a direct equivalence in the frequency domain; this is through relation (4.18).
3. It provides the properties of the residual process $\mathsf{y}_t - \mathsf{v}_t$ easily; this is due to (5.7) enabling (5.12) and (5.13).
4. Its analytical conveniences and properties are well-known; refer [84].

The optimal parameters using the measure in (4.9), with reference to (4.7) and (4.3), are given by

$$\widetilde{W}_t, \widetilde{\Gamma}^{\mathsf{z}}_h := \underset{W_t, \Gamma^{\mathsf{z}}_h}{\operatorname{argmin}}\ g. \tag{4.10}$$

Since orthogonal rotations of $W_j\, \forall j \in \mathbb{Z}$ lead to same $\Gamma^{\mathsf{v}}_h\, \forall h \in \mathbb{Z}$ in (4.7), there is no unique solution to the minimization problem (4.10).

Note 4.1. *In this thesis, the choice of the latent dimensionality q is made arbitrarily. No theoretical effort is spent towards the important problem of determining an optimal q. In the experiments, however, performance of the dynamic factor model across q will be evaluated. Asymptotic properties of the dynamic factors in the latent space with respect to larger measured number of samples and dimensionality is available in [37].*

The dynamic model defined by (4.1) - (4.8) implies that the measured vector random variable y_t is an addition of an independent linear process to another linear process formed by the dynamic transformation of a lower dimensional unit variance white noise. Apart from the time series aspect of the measured variables, the dynamic model bears much resemblance to the factor model: In (3.11), a latent vector random variable is transformed to an unobserved higher dimensional vector random variable which is perturbed by independent noise resulting in the measured vector random variable. This similarity invites the following definition:

Definition 4.2. *Let a q-variate latent zero mean unit variance white noise $\{\mathsf{x}_t\}$ be dynamically transformed by non-stochastic $\{W_t\}$ to an r-variate linear process $\{\mathsf{v}_t\}$. Suppose an independent r-variate linear process $\{\mathsf{z}_t\}$ is added to $\{\mathsf{v}_t\}$ to obtain an r-variate weakly stationary measured time series $\{\mathsf{y}_t\}$. Such a vector time series model which satisfies the conditions (4.1) - (4.8) and solves (4.10) is called a* ***dynamic factor model****.*

Model assumptions: Recall the original list of model assumptions in Section 1.2. With the dynamic factor model as per Definition 4.2, they may be concretely restated as follows:

1. the measured time series is a linear process,
2. the measured time series is a dynamic transformation of a zero-mean unit variance white noise of a dimensionality lower than that of the measured time series,
3. the *acvf* of the dynamically transformed process is a low rank approximation in a Frobenius norm sense of the measured *acvf*,
4. the residual time series is a linear process independent of the latent time series and has finite unique variances.

4.2 Spectral factor model

The objective of the dynamic factor model is to estimate the optimal parameters that maximize the commonalities of the measured time series $\{\mathsf{y}_t\}$ inherited by the unobserved time series $\{\mathsf{v}_t\}$. However, what stands out is the concern regarding how to perform such a maximization that adheres to the transformation of Γ_h^{v} to W_t as per (4.7).

Motivation for a Fourier-domain approach: It is clear from (4.7) that Γ_h^{v} is the correlation of the sequence $\{W_t\}$ in the time domain. According to the **autocorrelation theorem** of Fourier transform, which is also known as the Wiener-Khinchin-Einstein theorem, the autocorrelation of a function and power spectrum of that function are Fourier transform pairs; refer §10.1.1 in [77]. Then, $\forall h \in \mathbb{Z}, -\frac{1}{2} \leq \omega < \frac{1}{2}$

$$\Gamma_h^{\mathsf{v}} \overset{\mathcal{F}}{\longleftrightarrow} \mathsf{S}^{\mathsf{v}}(\omega) = \mathbf{W}(\omega)\mathbf{W}^*(\omega), \tag{4.11}$$

where $W_t \overset{\mathcal{F}}{\longleftrightarrow} \mathbf{W}(\omega)$ refers to the **discrete time Fourier transform** as per Definition 2.15 and $\mathsf{S}^{\mathsf{v}}(\omega)$ is the spectral density function of v whose (i,j)-th matrix element is $\mathsf{s}^{\mathsf{v}_i,\mathsf{v}_j}(\omega)$ $\forall h \in \mathbb{Z}, -\frac{1}{2} \leq \omega < \frac{1}{2}$. Note that applying the definition of the Fourier transform to the relation (4.3) gives

$$\mathsf{S}^{\mathsf{y}}(\omega) = \mathsf{S}^{\mathsf{v}}(\omega) + \mathsf{S}^{\mathsf{z}}(\omega). \tag{4.12}$$

It is further assumed that

$$\mathsf{rank}(\mathsf{S}^{\mathsf{z}}(\omega)) = r. \tag{4.13}$$

Also, it emerges from Property 2.7 that $|\mathsf{S}^{\mathsf{v}}(\omega)| = |\mathbf{W}(\omega)\mathbf{W}^*(\omega)| \neq 0 \quad \forall \omega \in [-\frac{1}{2}, \frac{1}{2}]$, i.e.,

$$|\mathbf{W}(\omega)| \neq 0 \quad \forall \omega \in [-\frac{1}{2}, \frac{1}{2}]. \tag{4.14}$$

For a finite τ-length realization of the process, combining (4.1) and (4.2) gives $y_t = \sum_{j=1}^{\tau} W_j x_{t-j} + z_t$. As per Definition 2.17, the **discrete Fourier transform** of these realizations are $\mathbf{y}(\omega_k) = \mathbf{W}(\omega_k)\mathbf{x}(\omega_k) + \mathbf{z}(\omega_k)$, at frequencies $-\frac{1}{2} \le \omega_k < \frac{1}{2}$, $k = 1, \ldots, \tau$.

To proceed, recall Theorem 2.5 where the spectral density function $\mathsf{S}^{\mathsf{y}}(\omega_j)$ becomes the covariance matrix of the **complex Gaussian distribution** of the discrete Fourier transform components sufficiently close to ω_j so that

$$\mathsf{y}(\omega_j) = \mathbf{W}(\omega_j)\mathsf{x}(\omega_j) + \mathsf{z}(\omega_j), \tag{4.15}$$

are complex Gaussian vector random variables $\{\mathsf{x}(\omega_j), \mathsf{y}(\omega_j), \mathsf{z}(\omega_j)\}$. Then, $\mathsf{x}(\omega_j)$ becomes a vector random variable whose complex Gaussian distribution has a covariance matrix

$$\mathsf{S}^{\mathsf{x}}(\omega_j) = I_q. \tag{4.16}$$

Note that the equivalent for (4.8) is

$$\mathsf{rank}(\mathsf{S}^{\mathsf{v}}(\omega)) = q. \tag{4.17}$$

Note that the form in (4.15) is very much reminiscent of the factor model, where $\mathsf{x}(\omega_j)$ is a latent factor of known Fourier characteristics transformed by a non-stochastic $\mathbf{W}(\omega_j)$ perturbed by independent vector random variable $\mathsf{z}(\omega_j)$. Hence, the motivation for pursuing a Fourier domain approach for the solution of the dynamic factor model is the possibility that classical factor model methods as reviewed in Chapter 3 might be availed to solve for W_t in (4.1).

Dynamic factor model equivalent in the Fourier-domain: Armed with a Gaussian probability distribution for the measured discrete Fourier transform $\mathbf{y}(\omega_j)$, the maximum likelihood estimation for the factor modeling should follow naturally. In that pursuit, the hope is to attain a relation connecting the maximum likelihood spectral density function $\mathsf{S}^{\mathsf{v}}(\omega_j)$, $\mathsf{S}^{\mathsf{z}}(\omega_j)$, and $\mathsf{S}^{\mathsf{y}}(\omega_j)$. Certainly, their inverse Fourier transform should yield their unique *acvf*s of Γ_h^{v}, Γ_h^{z}, and Γ_h^{y}, respectively, which are also of interest.
However, some estimate of the parameters of the interest is not satisfactory because the objective is to find those that will maximize the commonalities. Next, applying Theorem 2.1, the sum in (4.9) may be written as

$$g = \int_{-\frac{1}{2}}^{\frac{1}{2}} \|\mathsf{S}^{\mathsf{y}}(\omega) - \mathsf{S}^{\mathsf{v}}(\omega)\|_F^2 \, \mathsf{d}\omega. \tag{4.18}$$

Thereafter, in line with the arguments for (4.10), it could be deduced that the optimal parameters in the Fourier domain are

$$\widetilde{\mathbf{W}}(\omega), \widetilde{\mathsf{S}}^{\mathsf{z}}(\omega) := \underset{\mathbf{W}(\omega), \mathsf{S}^{\mathsf{z}}(\omega)}{\mathsf{argmin}} \; g. \tag{4.19}$$

Since orthogonal rotations of $\mathbf{W}(\omega)$ lead to same $\mathsf{S}^{\mathsf{v}}(\omega)$ in (4.11), there is no unique solution to the minimization problem in (4.19).

Due to the Fourier domain similarities of the dynamic factor model with the classical factor model, the following definition arrives:

Definition 4.3. *Let a q-variate latent zero mean unit variance discrete Fourier transform vector random variable* $\mathsf{x}(\omega_j)$ *be transformed by non-stochastic* $\mathbf{W}(\omega_j)$ *to an r-variate zero mean discrete Fourier transform vector random variable* $\mathsf{v}(\omega_j)$*. Suppose an r-variate discrete Fourier transform vector random variable* $\mathsf{z}(\omega_j)$ *that is independent of* $\mathsf{x}(\omega_j)$ *is added to* $\mathsf{v}(\omega_j)$ *resulting in an r-variate measured vector random variable* $\mathsf{y}(\omega_j)$*. Such a vector discrete Fourier transform model which satisfies the conditions (4.11) - (4.17) and solves (4.19) is called a* ***spectral factor model****.*

Model assumptions: Recall the list of model assumptions in Section 1.2 and subsequent to Definition 4.2. With the spectral factor model as per Definition 4.3, they may be restated as follows:

1. the measured discrete Fourier transform components ('spectra') are asymptotically Gaussian within small subbands,
2. the measured spectra are transformations of a zero-mean unit variance Gaussian spectra of lower dimensionality,
3. the spectral density function of the transformed spectra is a low rank approximation in a Frobenius norm sense of the measured spectra,
4. the residual spectra is a Gaussian independent of the latent spectra and has finite unique variances.

Basic goal of the spectral factor model: The dynamic and spectral factor models and the accompanying problem of maximization of the commonalities of a measured multivariate time series were defined in this chapter. The maximum commonalities transformation matrix is the best approximation, in a Frobenius norm sense, using a lower number of variables of the cross-covariances of the measured time series. Since there exist problems, as the examples in Section 1.2 show, where commonalities will aid learning, the goal is to adapt the transformation matrix for classification and prediction problems. This will be done by deriving the required parameters of the spectral factor model in Chapter 5 using the principle of maximum likelihood fostered by the constraint of maximum commonalities. Classification and prediction algorithms will be developed in Chapter 6.

4.3 Summary

In this chapter, the dynamic factor model and the spectral factor model were introduced. Conceptually, the dynamic and spectral factor models transform a latent vector random process by maximally inheriting the measured commonalities. It was discussed why the cross-covariances could be called as commonalities. A criterion based on approximating the *acvf*s in a Frobenius norm sense such that it will correspond to maximizing the commonalities was formulated. It was claimed that the inheritance of the commonalities of a vector random process by another increases if the Frobenius norm of the difference between their autocovariance functions across all lags decreases; an equivalent criterion for the spectral density function was also formulated. It was

assessed how the spectral factor model for measured discrete Fourier transform components in a ‘small’ bandwidth resembles the classical factor model.
The impediments of complex-valued parametric estimation should be overcame to extend the classical factor model estimation techniques reviewed in Chapter 3 to maximize the commonalities of the spectral factor model.