Chapter 8

Extensions

We had set out to build a model for multivariate time series that will enable developing strategies to predict or classify unseen data. In Chapter 4, the spectral factor model as the frequency-domain counterpart of the time-domain dynamic factor model was proposed. It was argued how dynamic transformation of a latent time series to a multivariate measured time series may be modeled to imbibe the characteristics that are common to any two measured variables. Those characteristics were called the commonalities. Maximally inheriting the maximum-likelihood cross-covariations was the estimation strategy adopted in Chapter 5 to model the commonalities. There, analytical formulae as well as an iterative algorithm for estimating the spectral factor model parameters corresponding to the optimal commonalities were presented. In Chapter 6, a classification rule was derived and a prediction methodology for multivariate time series based on the spectral factor model was designed. The experiments presented in Chapter 7 validate that the commonalities as defined, designed, and determined for the spectral factor model possess substantial classification and prediction capabilities for many real-world multivariate time series problems.
This thesis is concluded by highlighting a large number of possibilities that await in extending the spectral factor model. As pointed out, there are some improvements to the presented work that could be attained by overcoming the limitations and relaxing the assumptions.

The spectral factor model was developed aiming for applications involving multivariate time series learning. As seen in earlier chapters, the spectral factor model transformation matrix was the parameter that took the most focus in classification and prediction experiments that were carried out. There could be many more other applications possible through the spectral factor model along the same concepts as were presented. However, there are certain strong assumptions the spectral factor model is grounded on; they might pose some challenges for its widespread use. Therefore, in what follows, possibilities of developing the spectral factor model beyond the current design are investigated; some of the essential further investigations that are wished to be performed more formally are listed.

8.1 Challenges

Certain aspects that have come across as limitations and annoyances for the spectral factor model are as follow:

Linearity: The spectral factor model is rooted on the assumption that the measured multivariate time series is a linear weakly stationary process; this was essential for utilizing Theorem 2.5. In practice, given the samples of a realization of a time series, it is not easy to validate its linearity [13, 67]. This important, but broader scoped, issue was not addressed much in this thesis. The appropriateness of using Fourier domain methods is also at the mercy of this assumption. Hence, in using spectral factor model for the purposes dealt with in this thesis, it is recommended to tie the prediction or classification results with some appropriate test of nonlinearity of the measured multivariate time series data. Alternatively, lazy learning formalism of the spectral factor model could also be pursued whereby the model parameters will adapt to reflect the locality of the operating regime [16].
Commonalities: In developing the spectral factor model, existence of physically valid cross-covariation between the measured variables was naïvely assumed. It should, therefore, be borne in mind that applying it to independent or uncorrelated variables might show up numerically non-trivial off-diagonal *acvf* but whose interpretations might most certainly be illogical. In that respect, more robust estimation procedures of the sample spectral density of the type in [106] would be a path forward.
Pre-processing: As seen in the experiments, certain pre-processing of the time series was required to rectify obvious deviations of measured data from the assumption of weak stationarity. A rigorous and detailed study is envisaged for assessing the impact of the two preprocessing steps that used, viz., detrending and windowing [22, 65]; this is beyond the realms of this thesis.
Sample size: It is required a sufficient number of samples within a subband be maintained in order to obtain a reliable estimate of the spectral factor model parameters without inviting the curse of dimensionality [12]. Meanwhile, as dealt with in the experiments and in Algorithm 1, a sufficiently large number of subbands as required by Theorem 2.5 is to be maintained too. This balancing act was performed by testing on an array of choices regarding the number of parameters to be estimated and the sample size. This approach is perhaps not suitable when it is not clear whether the spectral factor model has learned. E.g., had the future series in the prediction exercise or true class labels in the classification experiments were lacking, evaluation of the prediction or classification accuracy, and therefore, the quality of the estimation would not have been possible. In essence, more theoretical efforts have to proceed beyond experimental validation and benchmarking towards determining an appropriate latent dimensionality q and the $\hat{\jmath}$ number of target frequencies for learning problems.

8.2 Further work

Certain realistic extensions to the spectral factor model beyond the objectives originally meant for it are listed below:
Process understanding: With an agreeable performance in a learning problem, it might be inferred that the presumed latent dimensionality q is credible. This hint regarding the dimensionality of the presumed latent process $\{\mathsf{x}_t\}$ is a preliminary step towards better understanding of a complicated high-dimensional time series generating system. In addition, there is a possibility to assert any process knowledge gained from human experience through the *acvf* Γ_h^{x} of the latent linear process. Such an *acvf* could replace the default assumption in (4.6) of $\{\mathsf{x}_t\}$ being a zero mean unit variance white noise.

Two other quantities that certainly will aid better understanding of the process under investigation are $\mathbf{x}(\omega_l)$ as per (5.30) and computed in (5.20) as well as $\tilde{\mathbf{v}}(\omega_l)$ estimated in (5.34). An inverse discrete Fourier transform of these quantities should enable now-casting [10], which is to provide a better assessment of the present and the past of the latent characteristics of the process.
Clustering: Suppose there are no class labels for an ensemble of various episodes from various time series processes and it is the intention to cluster them based on the characteristics of their commonalities. Then, a scheme similar to κ-means clustering [64] or any variations thereof might be adopted. Towards such a purpose, $\rho(i,k) = \sum_{j=1}^{\hat{\jmath}} \delta\left(\{\mathbf{W}(\omega_j)\}_i, \{\mathbf{W}(\omega_j)\}_k\right)$ may be used as the distance between any two time series episodes $\{y_t\}_i$ and $\{y_t\}_k$ computed across all the $\hat{\jmath}$ spectral factor model subbands where $\delta\left(\{\mathbf{W}(\omega_j)\}_i, \{\mathbf{W}(\omega_j)\}_k\right) = |\det(\{\mathbf{W}(\omega_j)\}_i^* \{\mathbf{W}(\omega_j)\}_k)|$ is the overlap for the j-th subband according to (6.1).
Real-time implementation: The computational aspects of a practical implementation of the spectral factor model was discussed in Section 6.1. There exist many multivariate monitoring applications, e.g., algorithmic trading [21], industrial plant monitoring [76], automated anesthesia [56], where frequent assessment and update of the model are necessary but prohibitive. In such problems, it is envisaged to use either the inexpensive EM-algorithm for incremental updates or approaches such as with online principal component analysis [73] for a real-time update to the analytical spectral factor model solution.

8.3 Summary

In everyday life, in business, health, search engines, etc., we are witnessing an immensely increasing demand for robust and efficient models for machine learning. Such models are necessary to meet objectives ranging from real-time computational decision support to scalable pattern recognition based on the multivariate time series they generate. It was demonstrated through reviews, contributions, and experiments that the dynamic and spectral factor models are live and active fields of research due to the simplified understanding of a complicated process they offer. The improvements and extensions to the modeling and learning frameworks of this thesis are near and feasible for them to deal with more diverse and real-world time series challenges. The spectral factor model promises to be a viable way forward for mastering the process generating the increasing volumes of multivariate streaming data.