Chapter 1

Introduction

The yearning to master a complicated process often prompts us to build its model. We wish to have a model that is simple but consistent with the characteristics of volumes of data obtained from the process. A model might serve several purposes: it could aid in representing and reviewing available data and to generate more data of similar characteristics. We also prefer the flexibility to evaluate the loyalty of the model to the characteristics of the given data and subsequently alter it if need be. Such a well-founded model should ultimately enable us to rein on the process.

The data typically comes as a set of samples and each sample is constituted by a set of **measured variables**. The characteristics of the measured variables might not be simple to comprehend. Hence, we wish the model to have a latent simplicity. To that end, we might conveniently demand the model to use a lower number of **latent variables** than the number of measured variables. Such a simplified interpretation of the process with a fewer number of underlying unobserved latent variables than the number of measured variables is called a **latent variable model** [11]. It is hoped that the data could be represented and reviewed with ease in terms of the latent variables.

In many applications, the set of measured variables of a sample is dependent on those of its preceding samples. The result is variation for a measured variable and covariation between the measured variables with respect to time and such data is called a **multivariate time series** [102]. The temporal variation-covariation across the measured variables of a multivariate time series is termed its **dynamic characteristics**.

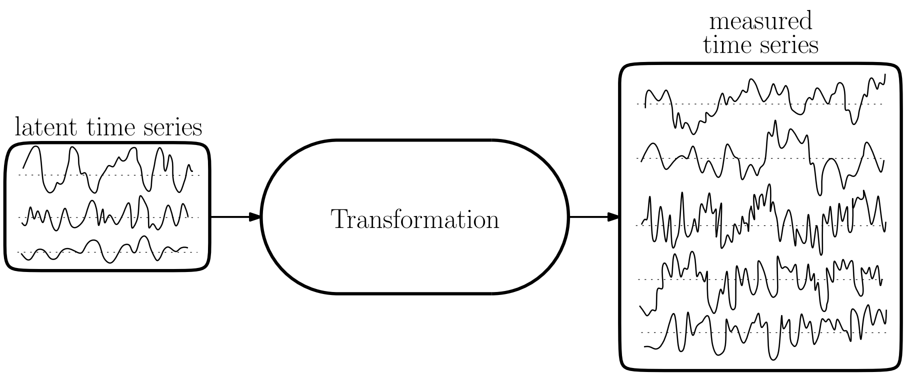

Figure 1.1: Three latent time series are dynamically transformed to five measured time series.

A model built on the dynamic characteristics of a time series is a reasonably flexible and accurate depiction of the underlying process.

Figure 1.1 illustrates a number of measured time series variables being generated by a transformation of a lower number of variables of the latent time series. It is this, possibly complicated, transformation that is to be modeled using the simplicity of the lower number of latent variables. Obviously, in doing so, we ought to be aware of the challenge that neither the transformation nor the number and characteristics of the latent variables are known. The modeling challenge is even greater because the lower number of variables might not be able to inherit the entire dynamic characteristics of the measured time series.

The scope of the problem of latent variable modeling of the dynamic characteristics of a multivariate time series is wide. It is natural, then, to restrict its scope as well as make it practically interesting. To this end, the dynamic characteristics common to any two measured variables [117] is deemed interesting for the modeling problem; those characteristics are termed the **commonalities**. In this thesis, commonalities will be defined in Definition 4.1 as the component cross-covariance functions of a weakly stationary multivariate time series.

For the still unknown model and latent time series, it is assumed that the latent variables are dynamically transformed to maximally inherit the commonalities of the measured time series according to some suitable metric. Thus, first we seek a **modeling framework** that defines the ingredients and scope of the latent variable model. The framework develops solutions for data-driven estimation of the **parameters** that control the transformation.

Apart from confirming our understanding of the process which generated the time series used to build the model, what could we do with a data-driven model? Suppose a model represents a collection of time series with similar commonalities. Then, the model could be used for **classification** of any new time series as belonging to that collection or not. This will be done based on similarities and dissimilarities of the commonalities of the new time series with those of the time series already available. Another utility could be in consistently generating future time series samples that bear characteristics similar to the time series we had used to build the model. Such a prospect might allow **prediction** of the time series given its past samples. It should be noticed that both of these applications involve applying the latent variable model towards unseen time series suggesting its ability to **learn**. Enabling the parameters of the transformation to predict and classify multivariate time series based on their commonalities implies a **learning framework** which is the main broad contribution of this thesis.

In Section 1.1 of this introductory chapter, a brief overview of the latent variable model with relevant references is given.

In Section 1.2, two practical examples to emphasize the motivation for a latent variable model with **dynamic transformation** are animated; these examples also form the experiments of the thesis. By these examples, interpretation of the commonalities in a multivariate time series as well as the learnability of a dynamic factor model are attempted. The basic assumptions that hold the latent variable model together and the basic strategy to arrive at a suitable model are listed. The technique of independent component analysis is complementary to the factor modeling pursued in this thesis; it is reviewed briefly.

In Section 1.3, the motivation for choosing the **dynamic factor model** as the latent variable model is stated. Its structure, it is discussed there, will transform the latent time series to measured time series dynamically while the transformation maximally inherits the commonalities from the measured time series according to some suitable metric. Using its frequency-domain counterpart called the **spectral factor model** for modeling and learning purposes is then vouched for.

In Section 1.4, the modus-operandii of learning a multivariate time series based on the commonalities is elaborated in layman terms. The modeling framework is introduced there; it is mentioned there how it is intended to bin discrete frequencies in subbands and maximize the inheritance via dynamic transformation of the measured commonalities within each of those subbands. The learning framework is also introduced there; the strategies for prediction and classification of multivariate time series using the spectral factor model are illustrated.

In Section 1.5, a brief review of existing works elsewhere in the growing literature of multivariate time series analysis that have similar objectives as that of the spectral factor model is conducted. Existing methodologies from diverse fields such as control systems, econometrics, biomedical signal processing, geology, etc., that are related to the ones used during various stages in the development of spectral factor model in this thesis are recapped there.

In Section 1.6, a pithy statement of objectives and in Section 1.7 a summary list of the main and supporting contributions of the thesis are provided. In Section 1.8, the organization of the thesis is outlined while in Section 1.9 a list of publications that motivated and aided this thesis are listed. In Section 1.10, a very essential summary of this notation-rich thesis is presented.

1.1 Latent variable model: An overview

A model of a process with a set of underlying variables held responsible for generating or representing a set of measured variables is the basic notion behind the latent variable model. Among the premiers to voice this notion loudly was Spearman in [114] and a series of works that followed. His research in psychology argued that there exists a statistical quantity called the 'general factor' that remains same in the scores of all mental tests on humans; whereas there is a 'unique factor' that varies with the tests. This idea has evolved over a century. Today it is all too common to conduct such tests where discrete responses to questions in the form of personality statements are assumed to be expressions of latent personality traits [121]. There also exist problems pursued in the sciences where it is necessary to assume that latent variables are not a continuum but discrete or categorical in nature. They are mainly of two types: First, the mixture model involves associating measured samples to a finite set of latent variables by estimating probabilities of the associations [82]. And the second type is the latent class model which pursues discrete latent variables that when presumed known or available amounts to locally independent measured variables [81]; this could be treated as a special case of the mixture model.

What qualifies as a latent variable model in this thesis in light of the above possibilities is the one which maps continuous latent variables to continuous measured variables. However, the important requirement stipulated is that there ought to be very few of the former in comparison to the latter. By this requirement, as envisioned by [55], the hope is "to attain scientific parsimony or economy of description." Considerable research has progressed in this arena known as **factor modeling** with an aim to explain correlations in the measured variables by a much lower number of latent variables: It was shown in [107] that three latent factors are generally sufficient in accounting for voltage variations, especially those relevant to electrocardiogram, recorded on the surface of a human body. In [69], yield rates of a large portfolio of stocks were shown to have a fewer number of latent factors corresponding to industry-wide common activities; whereas there were market factors unique to each of stock. Via factor analysis, six latent features out of twelve standard measured features were extracted for forecasting weather phenomenon in [9]. Factor models will be explored further in Chapter 3.

It will not escape our notice that in the seminal applications of factor models reviewed above, time dependency of the data was ignored for latent variable analysis. But this thesis focuses on the type of continuous latent variables that are to be modeled based on correlated samples in a multivariate time series. For the purpose of learning from such data, the classical factor model above will be insufficient and, instead, a **dynamic factor model** is required. Before entering into a detailed discussion on the dynamic factor model and its salient features in Section 1.3, the next section serves practical motivations for it.

1.2 Latent variable model: Two examples

In order to assert the context for a latent variable model for multivariate time series we discuss two practical examples below.

Classification of brain activities

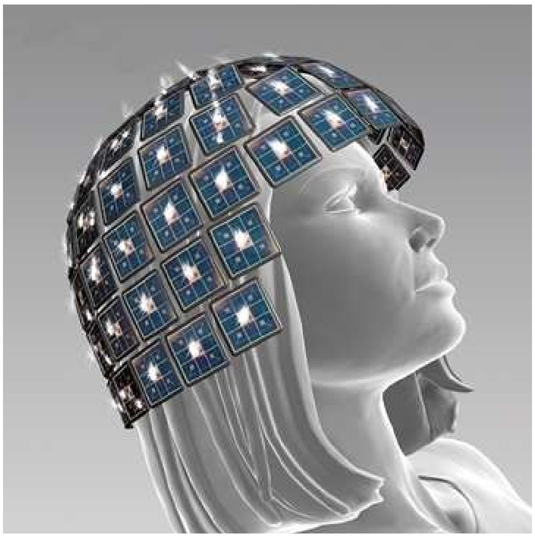

Figure 1.2: Illustration of the measuring of MEG signals via sensors positioned around the head [5].

Consider the scenario of a comfortably chaired computer gamer who makes smooth movements of a joystick by moving one of her wrists depending on the demands of a game. Under same experimental conditions, it might be assumed that activities in her brain are similar every time she makes the same wrist movement. Suppose we wish to do some experiments to know what could be going on in her brain for every wrist movement she makes. Remember that such experiments are very common these days and international conferences and competitions are conducted to learn more about brain activities [3]. The biggest beneficiaries of such studies include patients of neurological disorders [105, 115].

For the experiments, magnetoencephalography (MEG) signals from a human brain could be measured. These signals are based on magnetic fields induced by currents due to synchronized neuronal activities. Their recording is non-invasively performed via extremely sensitive magnetic sensors, as depicted in Figure 1.2; in reality, the sensors of an MEG scanner are encased in a well-isolated cavity in which the head is positioned comfortably. The signals have a temporal resolution of under a millisecond [42] and methods are available to attribute the readings from the sensors to designated spatial spots of the brain. Suppose ten signals attributed to ten spatial spots of the brain are measured. We know that these signals depend on one another mutually, i.e., activities in one part of the brain are influenced by activities in other parts. Figure 1.3 shows real signals from one such experiment [1]. We could perhaps observe various types of similar characteristics among any two measured MEG signals, i.e., delayed or inverted patterns, similar peaks and troughs but with one signal more fluctuating than other, etc. As a result, these signals could be considered temporally dependent on one another, i.e., current brain activity at a spot is influenced by current and previous activities at all spots.

For making a wrist movement based on some prompt, hypothesize the existence of only two **latent** activities in the brain of the gamer. This hypothesis could be based on a subjective opinion of an expert or mere guess. What they neurologically are is not relevant here. Nevertheless, assume that these two latent activities to be, e.g., (i) her cognition of the demands of the game and (ii) her reactions to move her wrist. In addition, suppose the general characteristics, e.g., averages, ranges, and other statistics, of these two fictitious latent signals of cognition and reaction are known.

The assumptions made so far are, firstly, the existence of a set of low-dimensional latent signals and, secondly, that their statistical characteristics are known. In addition, thirdly, assume that when the gamer has to make a particular wrist movement, the presumed latent cognition and reaction sequences undergo a particular transformation that gets expressed as mutual and temporal dependence seen in the ten measured signals. Although this assumption compounds to limiting the characteristics of the measured sequences as well. But it is a fair assumption because there ought to be a number of time dependent characteristics common to the ten MEG signals which are part of the same brain that collectively results in her making a particular type of movement of the joystick or another. For this reason, it is opined that the latent signals of cognition and reaction manifest themselves as the ten measured MEG signals consisting of a large amount of common variation-covariation, i.e., **commonalities**, corresponding to her brain activities. Then, essentially, **cross-correlations between the measured variables equals commonalities**. Obviously, there will be variations of the signals unaccounted by the commonalities, which will be unique or **idiosyncratic** characteristics pertaining to each of the measured signals and independent of the commonalities. Hence, the gamer making a wrist movement may be regarded as, the fourth in the list of model assumptions, that the latent signals transforming themselves maximally imparting desired commonalities to the measured signals according to some suitable metric; whereas, the fifth assumption is that any unaccounted variations of the measured signals are just undesired and independent noise.

Note 1.1. *From a data generation perspective, transformation of latent variables imparts commonalities to measured variables. From a modeling perspective, transformation of latent variables inherits commonalities from measured variables.*

To summarize, the model assumptions are

1. there exist generative latent variables of lower-dimensionality than the measured variables,
2. the statistical characteristics of the latent variables are known,
3. the transformation of the latent signals limit the modeled characteristics of the measured variables,
4. the transformation should maximally impart measured cross-correlation characteristics, and
5. the non-transformable characteristics are independent noise unique to each measured variable.

The presumed characteristics of the cognition and reaction latent variables stay same throughout the game; the gaming conditions will stay the same but challenges will differ. Then, it could be inferred that the common characteristics of the MEG signals during one wrist movement switch to a different class if and only if she changes the wrist movement to another class. This is a valid inference because one part of the brain behaves differently from another to various cognition and reaction challenges of the game she is playing. As a result, any class differences of the movement will manifest in the dynamic characteristics of the measured MEG signals. So a particular class of characteristics of the measured signals during a particular class of wrist movements is attributed to a corresponding class of transformation the latent variables undergo in imparting the commonalities.

The objective of this experiment is modeling multivariate time series for classification of wrist movements. Transformations corresponding to each cognition-reaction challenge are to be estimated and one class of transformations from another are to be distinguished. An approach could be to estimate, from all possible transformations, one that maximizes the **likelihood** to have generated the measured signals. Then, the estimate could be constrained further by requiring the presumed latent signals to maximally inherit, according to some suitable metric, the commonalities of the measured signals upon their transformation. It is now clear that the two steps:

1. estimate a maximum likelihood transformation based on model assumptions and
2. estimate the maximum likelihood transformation that inherits commonalities maximally as per a suitable metric.

Suppose we estimate the optimal latent variable model of the cognition-reaction process corresponding to each classified example of wrist movements. Then, as shown in Figure 1.4, for two classes of example measured MEG signals, we should be able to **classify** a test measured signal as belonging to a class of movements by computing how similar the commonalities of the test measured signal are to those in the classified examples. Obviously, the intrigue lies in classifying the measured signals without actually knowing or seeing the particular wrist movement she had performed.

Prediction of share prices

We take financial market as our next example where, suppose, the interest is in investing in a portfolio of shares of six companies, e.g., as shown in Figure 1.5, from various sectors of economic activities in a country. Suppose we know a successful investor who believes that investors are driven to purchase or sell shares based on perceived values of three underlying latent variables, viz., general political climate, consumer sentiments, and investor confidence. Of course, none of these fictitious latent variables could be metered objectively in practice. We wish to validate this belief before buying his advice. Note that as in the previous example, it is the number of latent variables and their presumed characteristics that is our concern and not their real physical or financial interpretations. If the investor's belief has merit, we could think of those latent variables to transform investment activities in the share market that manifest as changes in the share prices. Also, the latent variables when transformed must impart as much of the common dynamic characteristics, i.e., commonalities, demonstrated by the measured share prices.

In practice, even the best investors cannot consistently outsmart the market. And, our investor acquaintance above could blame any unexplainable fluctuations in the share prices on the dynamic characteristics of the share prices that the latent variables cannot inherit. These fluctuations could be **idiosyncratic** characteristics unique to each of those shares. However, if the transformation of the latent variable to the commonalities as we envisaged is true, we might be able to explain evolving tendencies of share prices. Therefore, in order to validate existence and influence of the commonalities, we could go by traditional investor wisdom to assess past behavior to bet on future: We could gather a **training** set of share prices of a sufficiently long evolution of various shares of the portfolio. We could then then estimate a **dynamic transformation** that is optimal in the sense of having the **maximum likelihood** to have generated the training series. Subsequently, we could search among the maximum likelihood transformations one that will **maximally inherit**, according to some suitable metric, the commonalities of the share price evolution process. We could use a predictor that is based on minimizing temporal tendencies to err in predicting the training series. The set of parameters of such a predictor will be a function of the optimal dynamic transformation. Then, given a current evolution of the share prices, it should be possible to **predict** their future evolution with a reasonable accuracy.

Independent components versus latent factors

The thesis, as discussed so far, involves estimating a generative model where a set of latent variables are transformed to a larger number of measured variables based on the latter's characteristics. To estimate the transformation matrix, the maximal inheritance of the mutually dependent variation-covariation characteristics was the criterion considered.

In a complementary setting, there exists a wide body of literature called independent components analysis or blind source separation [27, 24]. Independent component analysis is often called 'non-Gaussian factor analysis' [61]. In contrast to the objective of factor analysis, the objective in independent components analysis is to identify mutually 'independent' latent variables.

One of its working philosophy is due to the central limit theorem whereby any transformation of the latent variables will be maximally non-Gaussian if it equals one of the independent latent variables; hence, latent variables are considered non-Gaussian [60]. In contrast, factor analysis stresses on dependencies and Gaussians are readily accepted as the latent variables.

In another working philosophy of the independent components analysis, higher predictability of a latent series component than that of any dynamic transformation of the latent series components is exploited to sequentially identify the latent variables [26]. In dynamic factor analysis as presented in this thesis, higher cross-correlations via commonalities aid predictability. On the other hand, in this thesis, the variation-covariation characteristics of the latent variables, their mutual dependence or independence, will be assumed known.

Moreover, in this thesis, the transformation of the latent variables will be assumed a linear process; therefore, the measured variables are also assumed linear processes. The focus in this thesis is in estimating a transformation for the latent variables rather than identifying the latent variables themselves as done in a blind source separation problem.

1.3 Dynamic and spectral factor models

As the two examples above highlight, the processes that are of our interest generate data samples such that each measured variable is free to influence the preceding samples of itself and other variables. This emphasizes that the order in the sequence of occurrences of the measured samples is rather important and it must be indexed appropriately. It is convenient to attribute the index of the sequence to discrete instants of time. This is the reason we call such a sequential collection of correlated data samples a **time series**.

In many processes we measure a set of variables at the same instant. This implies that every sample of the data is formed by the same ordered set of multiple variables. Such a collection of samples is referred to as **multivariate** data.

This thesis focuses on learning from **multivariate time series** where any measured variable in a data sample is influenced by, in general, the rest of the variables in the sample and all the variables of all the preceding samples. Such an influence could be quantified as a function of the **lag**, which is the number of time instants by which two samples differ. So, when a multivariate time series is said to display dynamic characteristics, the term **dynamic** attributes its characteristics to be **lag-dependent**.

In the context of multivariate time series, the driving assumption is that a lower number of latent variables are transformed to a number of measured variables resulting in a latent variable model as illustrated in Figure 1.1. A practical motivation for that assumption is that a fewer number of variables will aid simplicity in interpretation, modeling, and computation. Note, however, that the true latent variable transformation is unknown and estimating it is part of the objective of this thesis.

Recall that the characteristics of a measured time series are to be modeled. But how could simplicity in modeling be aided when unknowns such as latent time series and variable transformation are injected into the model? In that respect, either or both the transformation and the characteristics of the latent time series could be assumed unknown. Remember, we wish to strictly control the underlying process which the latent time series represents and prefer it to have characteristics not as complicated as those of the measured time series. Moreover, if possible, expert opinion on the latent time series could be invited. Hence, it will be assumed that the latent variable characteristics are known and the transformation is unknown.

To enhance simplicity even further, the latent variables will be assumed a multivariate time series with lag-independent characteristics whereas it is the transformation that is dynamic and unknown. The challenge then is to estimate the dynamic transformation that best generates the measured time series from the latent time series. In this framework, the latent variables upon transformation are assumed to impart the dynamic characteristics to the measured time series. Hence, given a dataset of measured time series, such a framework implies estimating the ideal transformation that could yield the desired dynamic characteristics. This is illustrated in Figure 1.6, where the 'desired time series' is enabled to capture the desired dynamic characteristics pertaining to the measured time series; whereas the 'undesired time series' is the difference between the measured time series and the desired time series. The set of parameters θ of the dynamic transformation are retained for reference.

Note 1.2. *Figure 1.1 depicts the unknown true transformation that generates the measured time series from a latent time series of unknown characteristics. Whereas an appropriate dynamic transformation of Figure 1.6 has to be estimated based on the measured time series and the presumed characteristics of the latent time series.*

Remember that the desired dynamic characteristics of the measured time series are its commonalities. As introduced earlier and through the examples, maximally capturing the commonalities is tantamount to learning. It has been decided to keep the latent time series characteristics known, lag-independent, and simple; they are the underlying factors of the model. The model which consults the measured time series to dynamically transform the factors to maximize the commonalities is named the **dynamic factor model**. This concept is illustrated in Figure 1.7, where the term **idiosyncrasies** refers to the undesired time series that retains no commonalities. Hence, a dynamic factor model is a multivariate time series model which dynamically transforms a latent time series of predetermined characteristics to maximally, in some suitable metric sense, inherit the common dynamic characteristics of a set of measured multivariate time series. It accepts measured time series as input and outputs commonalities, idiosyncrasies, and the optimal model parameters.

One possible dynamic characteristic of the measured variables is periodicity. There could be many periodic dynamic characteristics in the measured time series. A periodicity corresponds to a **frequency**, which is associated with the number of time series samples that constitutes the period. Decomposing the measured time series into component frequencies is intuitively simple, analytically rich, and practically useful. Such a decomposition of a time series across all possible frequencies is called the **spectral**

analysis [99]. An inverse synthesis of frequency components to time-domain is also possible through spectral analysis. This is a motivation to understand the influence of various frequencies in the dynamic characteristics of the measured time series. As depicted in Figure 1.8, such a frequency spectral analysis of dynamic factor model would require analyzing measured and latent time series, commonalities and idiosyncrasies, and dynamic transformation all in the spectral or frequency-domain. It will be called the **spectral factor model**, which may be contrasted with the time-domain equivalent in Figure 1.7. In that respect, Figure 1.9 depicts the frequency spectral equivalent of the dynamic factor model. Note that Figure 1.9 has the same input and outputs as the dynamic factor model in Figure 1.7 for they are subjected to spectral analysis and its inverse, respectively.

1.4 Learning by maximizing spectral commonalities

The appeal of the frequency-domain approach in many fields of study are mainly due to the computational advantages and the physical interpretation it offers [97, 23]. Many time-domain processing requirements of a time series may be easily realized in the frequency-domain; the software and hardware implementation of such processing is widely available [63]. These further motivate, in addition to the theoretical appeal, the development of a spectral factor model for learning from multivariate time series.

The spectral components correspond to an infinite continuum of frequencies, but samples from a discrete time series are practically limited. This limits and motivates targeting just a set of discrete frequencies. But uncertainty is encountered in balancing resolution and precision of the spectral components at these discrete frequencies. To tackle the challenge, spectral components in small non-overlapping bands of frequencies may be considered. In these **frequency subbands**, spectral factor modeling might be performed by assigning probabilities to various discrete spectral components of the measured time series. The aim is to estimate a probabilistic spectral factor model that is the most likely to affiliate the measured spectral components. For this purpose, model parameters that will maximize the likelihood of simultaneous occurrences of all the measured spectral components within a subband will be probed. From all possible maximum likelihood spectral factor models, the one which maximally, in some suitable metric sense, inherits the measured commonalities on the dynamically transformed factors could be chosen. Recall that commonalities are cross-correlations of the measured variables. Later in the thesis, their inheritance by the dynamic factor transformation will be defined as a very simple and intuitive function of **all cross-correlations of the measured variables over all lags**.

Figure 1.10 illustrates the strategy for **maximum likelihood maximum commonalities** spectral factor model estimation. The spectral components of the presumed latent spectra and the given measured spectra are divided into frequency subbands. For each subband, maximum likelihood estimation of parameters of the spectral factor model will be performed. Two distinct maximum likelihood estimation methods will be demonstrated: The first method is an **analytical estimation** which gives an explicit formula for the optimal parameters. The second method is an **iterative estimation** starting with initial guesses of the model parameters that are updated till they converge to possible optimal parameters. Further, for each of those methods, techniques to extract those parameters that will maximize the commonalities are devised.

Commonalities of the measured time series maximally inherited in some suitable metric sense by the dynamic factors allow the model to learn a process. **Classification** of multivariate time series measured from various distinct processes is the first of our two learning applications. In Section 1.2, the example of classification of MEG signals involved in maneuvering a joystick via wrist movements was discussed in detail. The various classes there could be regarded as dynamically transformed latent signals corresponding to various visual prompts on a computer monitor. There will be several example time series in a class. For each such example dataset obtained from any two processes deemed to be distinct by an expert, two classes of spectral factor model examples are built. The models are considered to have learned the example processes upon maximally inheriting their commonalities on their respective maximum likelihood parameters according to some suitable metric. Then, in order to decide which of any two possible processes a new unclassified measured time series belongs to, the commonalities of the new dataset need to compared with those of the two classes of spectral factor models. Based on the discussions so far, the commonalities will determine the dynamic transformation. In that regard, the new test measured time series will be assigned to the class to which its estimated dynamic transformation has the most proximity to. Such a strategy for the classification exercise requires a comparator of the dynamic transformations as shown in Figure 1.11. This method could be extended to associate a time series as belonging to one of any number of identified classes of processes.

Prediction of multivariate time series is chosen as the other learning application. Once knowledge of the characteristics and the number of latent time series variables are presumed, a spectral factor model based on a training set of measured time series could be estimated. Based on the optimal dynamic transformation that maximally, according to some suitable metric, inherits commonalities of the measured time series, a multivariate time series predictor could be built . The example of a portfolio of share prices that was discussed in Section 1.2 is used for prediction experiments later in the thesis. For a given length of training time series, a number of latent time series less than the number of the measured time series are experimented with to build the spectral factor models. Using their parameters, a prediction framework based on minimizing the prediction error given past samples is built. As shown in Figure 1.12, a future evolution could be charted for a given current evolution. The prediction accuracy will be validated using the true share prices whenever it becomes available.

1.5 Dynamic and spectral factor models in literature: A brief review

It must be mentioned at this juncture that the concept of commonalities and dynamic factor model is not very new. In one of the earliest formal studies about dynamic factor model, its estimation in the Fourier domain was famously attempted by [100] for advanced control systems and [104] for macroeconomic forecasting. An idea similar to commonalities was promoted by [104] in econometrics literature as "common shocks." Like in this thesis, they too state the relation between the spectral density functions via a likelihood function of the discrete Fourier transform components within disjoint frequency subbands. Then, they obtain maximum likelihood parameter estimates via Fletcher-Powell optimizations and standard hypothesis testing procedures. However, they stop short of going much farther than the possibility of infinitely many uncon-

strained solutions for the spectral factor transformation matrix.
Another line of approach is an approximate dynamic factor model with finite lags as was developed in [118]. There, the estimation was performed as the principal components of an expanded set of static factors; their aim was prediction of macroeconomic variables. Their prediction equations take the form of vector autoregression where the estimated static factor components may be directly plugged in without having to estimate the Fourier domain parameters as in fully dynamic models.
Recently, [36] changed the landscape of research in this domain substantially with their generalized dynamic factor model, e.g., they spell out the extent of flexibility allowed for idiosyncrasies and derived the convergence properties of the model parameters as the number of samples and measured variables grow. They focus on forecasting macroeconomic variables and the work forms a series of highly acclaimed and rigorous treatment of the subject. There are agreements between the parts of the approach to the problem in this thesis and theirs in (i) concluding that the principal components of the spectral density matrix gives the analytical solution (ii) the idiosyncrasies could be mildly cross-correlated. However, the ideas introduced in this thesis are quite different from theirs; e.g., an iterative estimation procedure and a time series classification strategy are provided. Moreover, while this thesis focuses in the multivariate time series modeling and learning frameworks, they focus on prediction of latent commonalities.
In §7.8 of [111] a maximum likelihood estimation estimation of dynamic factor model in the spectral domain much like in this thesis is pursued. They use it for analyzing function magnetic resonance imaging data. Their final analytical solution overlaps with the one developed in this thesis and in [36]. But they seem not to share any qualms regarding the non-analytical nature of the log-likelihood function and does not see such a model from the classification or prediction perspectives. They do not provide an iterative solution strategy either.
Among the front-runners of the dynamic factor model was [90] who wanted to estimate the latent trajectory of a patient's state based on vital signals. He rewrote the dynamic factor model parameters as a Markovian state model whose estimation was carried out via Kalman filter principles.
Now, let us divert the attention to a spectral domain method whose priority was multivariate time series classification rather than prediction. In [66], sample spectral densities are compared for classifying and clustering episodes of multivariate time series. Their experiments involved discriminating between time series generated by earthquakes and those by explosions. However, they do not consider existence of a low-dimensional latent time series and, as a result, were able to design disparity measures that work by comparing the full-rank sample spectral densities. This thesis uses the information contained in a rank-deficient maximum-likelihood maximum-commonalities spectral factor transformation matrix to perform classification.

1.6 Objectives

Based on discussions on the motivation and the premise of this thesis so far, its objectives are broadly divided into developing

1. *a multivariate time series latent variable modeling framework*
   To meet this objective, dynamic and spectral factor models as well as commonalities are formally introduced and defined in Chapter 4. The maximum likelihood

maximum commonalities spectral factor model is derived in Chapter 5. An analytical form as well as an iterative procedure for estimating such a spectral factor model are developed.

2. *a multivariate time series maximum commonalities learning framework*
   This objective is achieved by providing multivariate time series classification and prediction algorithms in Chapter 6, which exploit the maximum commonalities parameters of the spectral factor model.

1.7 Contributions

The following is the list of main contributions of this thesis:

- ▷ The most original contribution of this thesis is the development of a commonalities-based classification metric in (6.4) that compares overlap of spectral factor model subspaces to distinguish multivariate time series processes.
- ▷ The second most important contribution is the utilization of the estimated commonalities in developing a multivariate time series prediction strategy via classical vector autoregression on current and past samples; it is detailed in Section 6.3.

The following is the list of supporting contributions of this thesis, which are improvements, interpretations, or alternatives to existing work in the literature:

- ▷ Derived an analytical solution for spectral factor model in (5.10) using low-rank approximation theorem.
- ▷ Derived an iterative solution for spectral factor model in Section 5.2 using the Expectation - Maximization algorithm whose converged parameters that maximally inherit the commonalities are extracted by applying the Gauss - Markov theorem in Section 5.2.3.
- ▷ Obtained the mild cross-correlation property of the idiosyncrasies in Property 5.1 via Weyl's theorem.
- ▷ Used Wirtinger relaxations for maximizing log-likelihood in Chapter 5.

1.8 Organization

A non-technical overview of the thesis was presented so far. In the two chapters that follow, the basics on which this thesis is built is presented.

- In Chapter 2, an essential overview of multivariate time series analysis is provided; very essential time-domain and frequency-domain analyses are presented there.
- In Chapter 3, parametric estimation methods for probabilistic models concisely and as required is discussed.

With much groundwork done with the aforementioned chapters, the two chapters that follow introduce and develop the dynamic factor model framework to suit the learning framework objective of this thesis.

- In Chapter 4, a technical introduction and motivation for the concepts of dynamic and spectral factor models as well as commonalities and their maximization are provided.
- In Chapter 5, an analytical method and an iterative method for maximum likelihood maximum commonalities spectral factor model are derived.

Subsequent to the development of the dynamic factor model, the learning framework is provided.

- In Chapter 6, a time series learning framework is built using the inherited commonalities by explicitly stating algorithms for classification and prediction of multivariate time series analysis.

The contributions are tested and possible extensions are discussed in the last two chapters:

- In Chapter 7, the methodology and results of multivariate time series classification and prediction experiments are presented.
- In Chapter 8, improvements and plans for further research and applications are mentioned.

1.9 List of relevant publications

(I) Miranda, A. A., Olsen, C., Bontempi, G.: Fourier spectral factor model for prediction of multidimensional signals, *Signal Processing*, 91(9):2172-2177, Elsevier, 2011.
This paper presents the vector autoregressive prediction of a multivariate measured time series on the current and past samples as developed in Section 6.3 using the autocovariance of the maximally inherited commonalities. It demonstrates prediction of yield rate of a six-variate share portfolio with substantially better accuracy than standard vector autoregression; the daily prices of those yield rates are used for experiments in this thesis.

(II) Miranda, A. A., Bontempi, G., Schuddinck, P.: Fourier spectral factor model for classification of high-dimensional MEG signals, *Under review in Biomedical Signal Processing and Control*, Elsevier, 2011.
This paper presents the commonalities-driven classification strategy developed in Section 6.2 for multivariate time series; the magnetoencephalography experiments conducted for Section 7.1 are also presented.

(III) Miranda A. A, Caelen O., Bontempi, G.: Machine learning for automated polyp detection in computed tomography colonography, *Biomedical Image Analysis and Machine Learning Technologies*, Medical Information Science Reference, 2009.
This paper compares a number of classifiers well-known in machine learning that perform well despite severe imbalance in the class representation and unreliable features. That classification problem may be compared with that in Section 6.2 to understand that popular robust classifiers designed towards identically and independently distributed data are not directly usable for a multivariate time series classification problem.

(IV) Miranda, A. A., Le Borgne, Y.-A., Bontempi, G.: New routes from minimal approximation error to principal components, *Neural Processing Letters*, 27(3):197-207, Springer, 2008.
This well-cited paper discusses the classical principal components analysis from a layman perspective. Principal subspaces, eigenvalue decomposition, trace minimization are recurrent themes in this thesis and are presented in simple terms in the paper.

(V) Miranda, A. A., Whelan, P. F.: Fukunaga-Koontz transform for small sample size problems, *Proceedings of the IEE Irish Signals and Systems Conference*, pp. 156-161, Dublin (2005)
This paper discusses a strategy for comparing the principal subspaces due to the autocorrelation matrices of two classes of multivariate data in a common full-rank space. The features of this paper such as real-valued projections, euclidean distance measures, binary classification, etc., are serious shortcomings for comparing multiple spectral factor subspaces and to overcome them the classification metric in (6.4) was developed.

1.10 Notations

Herein, notations and conventions used in this thesis are introduced. Unfortunatelely, terms whose proper definitions will show up in later chapters only will be mentioned here. Nevertheless, it is important to read this section carefully for grasping the treatment of technical aspects later.

The following convention of using Latin characters is adhered to: Incremental variables such as indices are denoted using i, j, k, and l.

> **Note 1.3.** *From Chapter 4 onwards, certain alphabets are appointed to imply the same variable for the rest of the thesis. These are, respectively, q and r for the latent dimensionality and observed dimensionality. The letters v, x, y, and z are used for transformed, latent, measured, and idiosyncratic variables; but they will have an appropriate meaning depending on whether it appears in Roman, sans-serif or boldface fonts.*

Use of t for time indices and h for time delays are reserved throughout. The aforementioned conventions imply that both scalars and vectors are denoted in small-case. Linear algebra drives much of the contributions and a rectangular matrix is always in capital-case as in X.

Ideas from the basics of probability and stochastic processes are used liberally. A sans-serif font such as in x is used to denote a random variable and its realization will be in Roman font as in x.

> **Note 1.4.** *Random variables and vector random variables, either real-valued or 'complex-valued', will be denoted in the same fashion using a sans-serif font; the context will make their distinction clear.*

Also, the sans-serif font will be used to denote common mathematical operations or functions such as log for natural logarithm, p for a probability density function, S for spectral density function, etc.

The standard practice of using a blackboard bold font to denote number sets, e.g. set of complex numbers $\mathbb{C}$, set of integers $\mathbb{Z}$, etc. are followed. However, a calligraphic font will be used to denote a group of items such as two classes $\mathcal{C}_1$ and $\mathcal{C}_2$ and the Gaussian family of probability densities $\mathcal{N}$.

Certain Greek alphabets will denote the same variable, function, or metric throughout, e.g., μ for mean and Γ for autocovariance function matrix.

Subscripts are used for indices in two capacities: First, they denote indices as in x_t for the t-th time sample of x. Second, they denote a component of a vector or a matrix. E.g., x_k is the k-th component of random variable x and X_{ij} is the i, j-th element of matrix X. This gives the possibility to interpret nested subscripts appropriately. E.g., x_{k_t} is the t-th time sample of x_k and the inner subscript, i.e., k in x_{k_t}, will be always interpreted as the component index and the outer subscript, i.e., t in x_{k_t}, as the sequence index.

> **Note 1.5.** *Other than its usual interpretation as scalar exponent, superscript on a function or an operator will denote the operand, e.g., μ^{y} denotes mean of the random variable* y.

Fourier analysis is a persistent theme in this thesis and boldface, e.g., $\mathbf{x}(\omega_k)$, implies discrete Fourier transform components.

Presented below is a table of certain frequently used symbols and notations:

u', U'	transpose of vector u or matrix U
$\bar{u}$, $\overline{U}$	complex conjugate of scalar or vector u or matrix U
u^*, U^*	complex conjugate transpose of scalar or vector u or matrix U
$\lvert u\rvert$, $\lvert\mathcal{U}\rvert$	absolute value of scalar u; cardinality of set $\mathcal{U}$
$U_{i:j}$	matrix formed by columns $i, i+1, \ldots, j-1, j$ of matrix U
$\mathrm{det}(U)$	determinant of real or complex-valued square matrix U
$\{\mathsf{u}_t\}$	time series due to sequence of random variables u_t $\forall t \in \mathbb{Z}$
u_t	realization of a time series $\{\mathsf{u}_t\}$ at instant t
$\mathbf{u}(\omega_j)$	discrete Fourier transform due to $\{u_t\}$ at frequency ω_j
$\mathsf{P}(\mathcal{U})$	probability of the event $\mathcal{U}$
p^{u}	probability density of (a possibly vector) random variable u
p	order of vector autoregression
E^{u}	expectation with respect to p^{u}
μ^{u}	mean of (vector) random variable u
Γ^{u}	variance (covariance matrix) of (vector) random variable u
$\Gamma^{\mathsf{u},\mathsf{v}}$	cross-covariance (matrix) of (vector) random variables u and v
acvf	autocovariance function
γ_h^{u}	*acvf* of univariate $\{\mathsf{u}_t\}$ at lag h
Γ_h^{u}	*acvf* of (univariate or multivariate) $\{\mathsf{u}_t\}$ at lag h
i	imaginary operator
I_q	identity matrix of size $q \times q$
$\widetilde{\mathrm{diag}}(U)$	setting off-diagonal elements of U to zero
$\lVert U\rVert_F$	Frobenius norm of matrix U
$\mathcal{F}$	Fourier transformation; discrete Fourier transform
iid	independently and identically distributed
$\langle x\rangle$	*A posteriori* mean of x
L	log-likelihood function
$\mathcal{D}$	a dataset
τ	length of a time series realization
$\hat{\jmath}$	number of frequency subbands
κ	number of relevant nearest neighbors
W	dynamic factor transformation matrix
$\mathbf{W}$	spectral factor transformation matrix
$\widehat{\mathbf{W}}$	maximum likelihood $\mathbf{W}$
$\widetilde{\mathbf{W}}$	maximum commonalities $\widehat{\mathbf{W}}$
S^{u}	spectral density function of $\{\mathsf{u}_t\}$
$\check{\mathsf{S}}^{\mathsf{u}}$	sample spectral density function of $\{\mathsf{u}_t\}$
$\widehat{\mathsf{S}}^{\mathsf{u}}$	maximum likelihood S^{u}
$\widetilde{\mathsf{S}}^{\mathsf{u}}$	maximum commonalities $\widehat{\mathsf{S}}^{\mathsf{u}}$

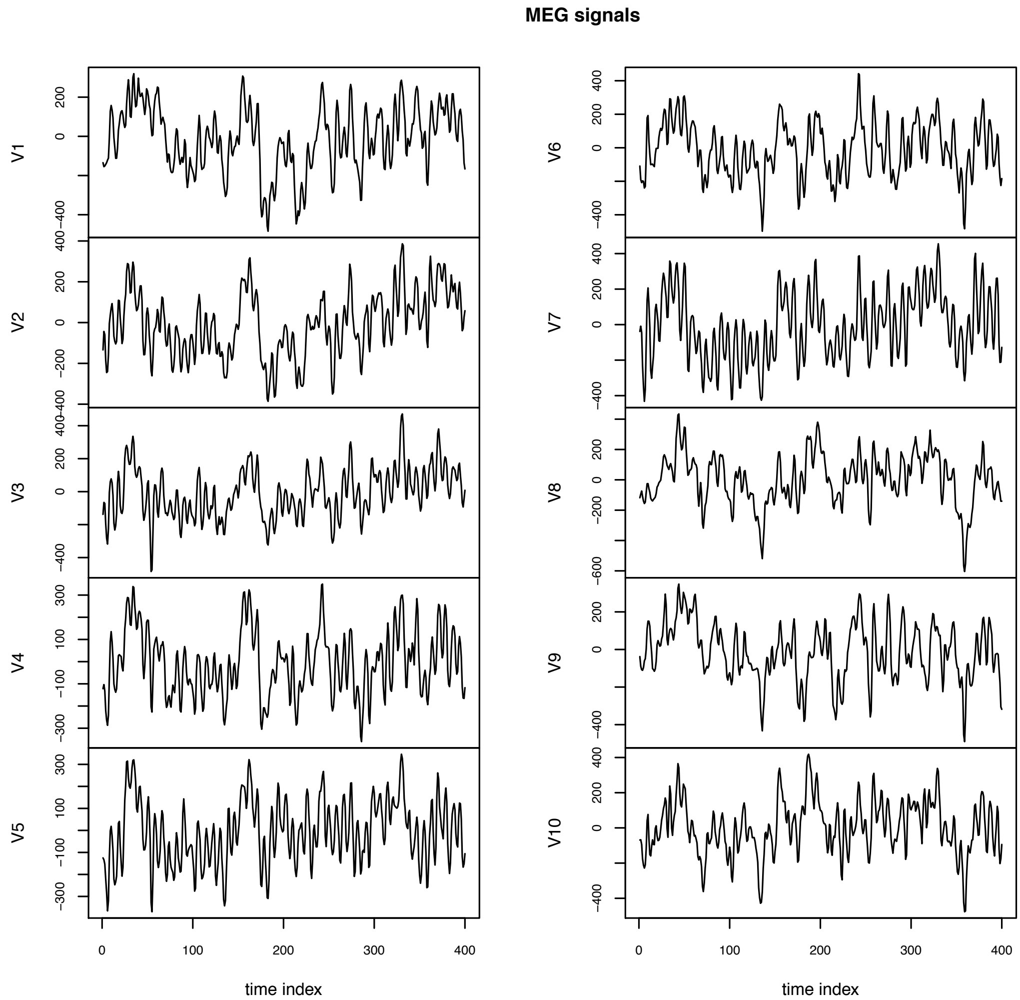

Figure 1.3: MEG signals corresponding to ten spatial spots V1-V10 of the brain upon a particular movement of the wrist.

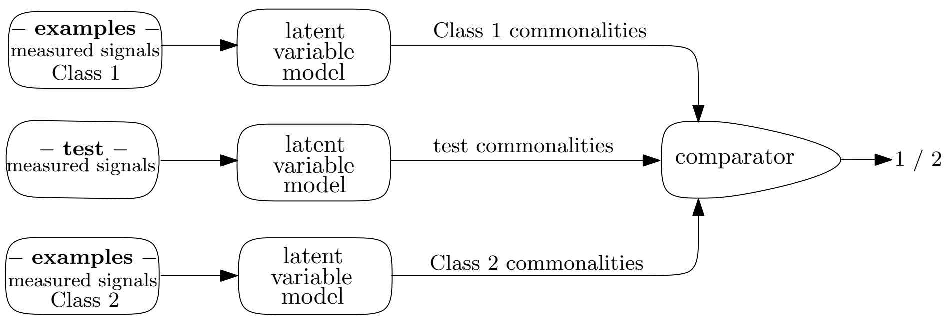

Figure 1.4: Among the two classes, a test measured signal is associated to the one to which its commonalities are closest to.

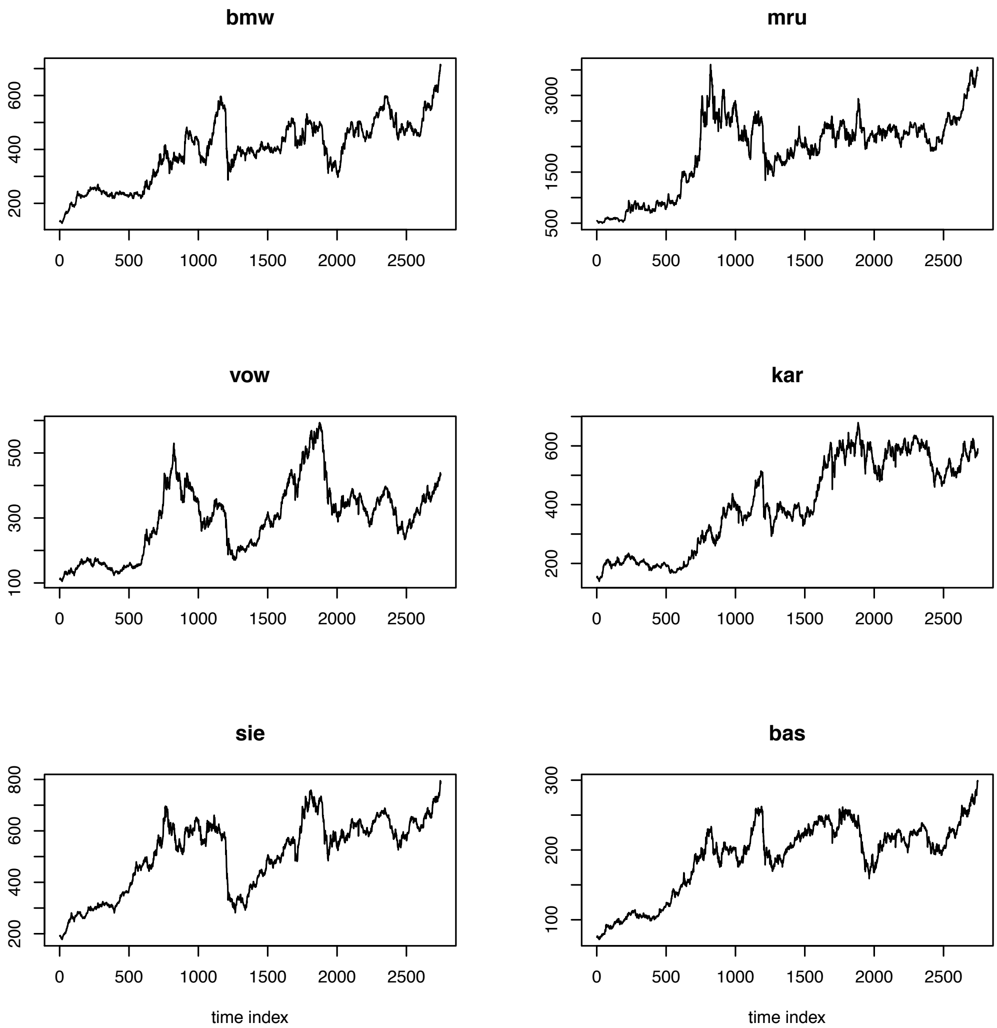

Figure 1.5: Daily stock prices in Deutsche Mark of six German companies between 01/01/1983 - 30/12/1993 [6].

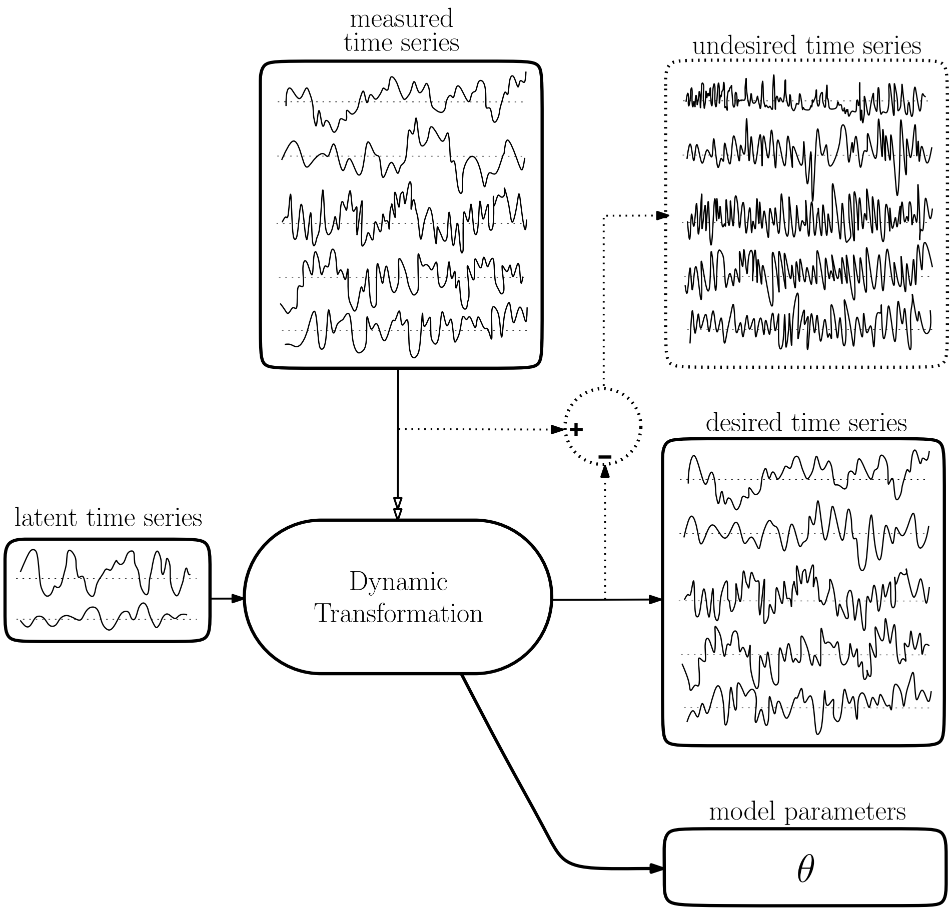

Figure 1.6: Dynamic transformation, whose parameters are summarized by θ, of the latent time series will consult the measured time series to decompose the latter into desired and undesired time series.

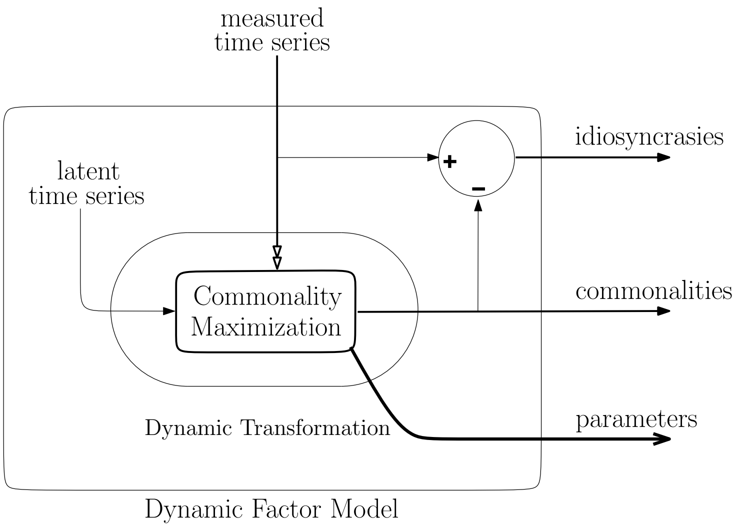

Figure 1.7: According to some suitable metric, the dynamic factor model allows the dynamic transformation to maximally inherit the commonalities from the measured time series; their difference forms the idiosyncratic time series.

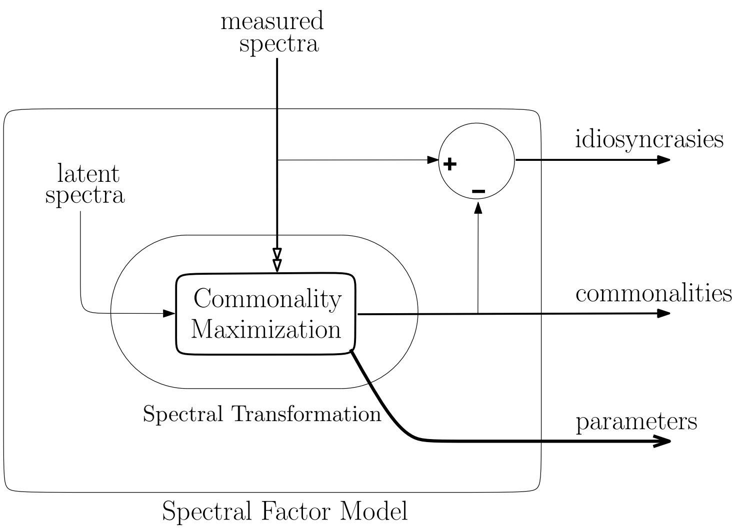

Figure 1.8: The spectral factor model expresses dynamic factor model in frequency-domain.

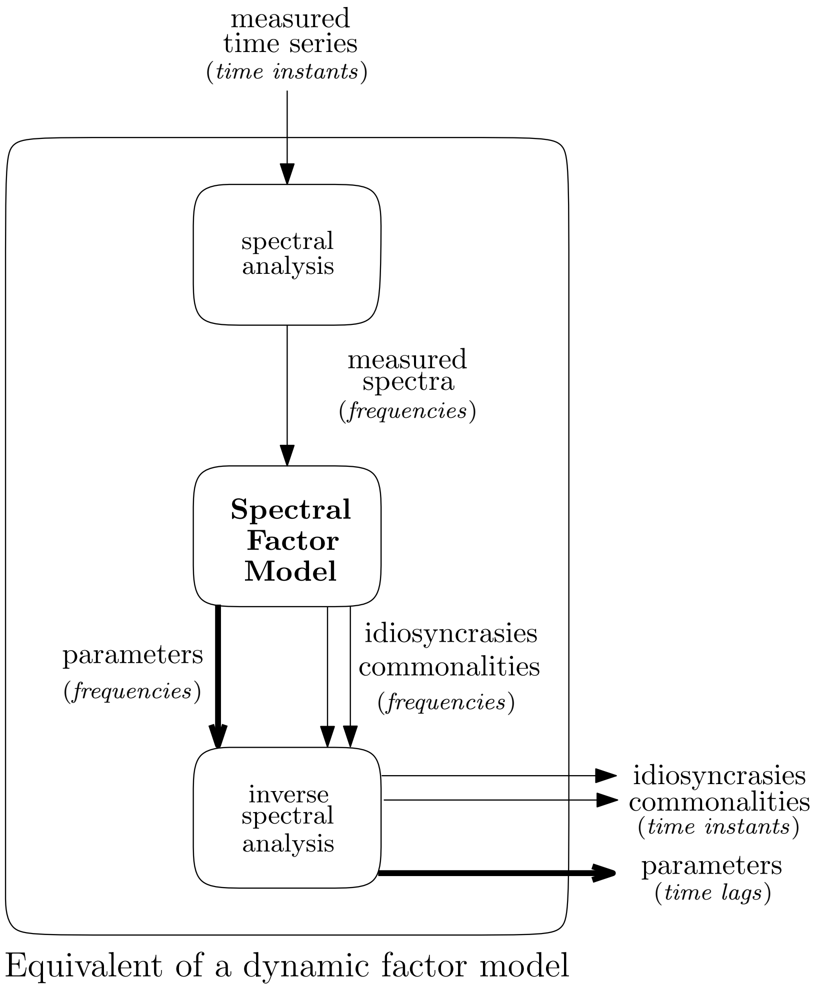

Figure 1.9: An equivalent of the dynamic factor model is built by sandwiching the spectral factor model between spectral analysis and its inverse operations.

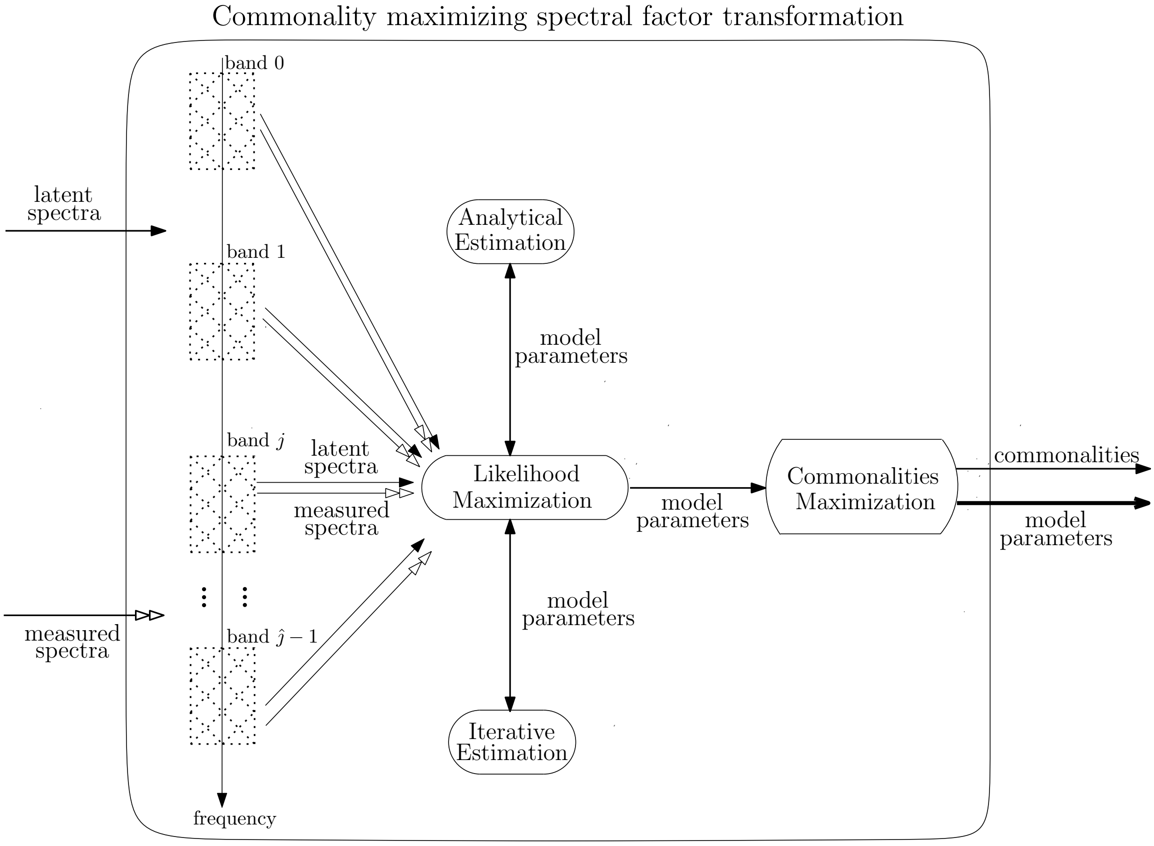

Figure 1.10: Maximized commonalities for a finite $\hat{\jmath}$ number of individual frequency bands are obtained from amongst the family of maximum likelihood spectral factor model parameters analytically and iteratively.

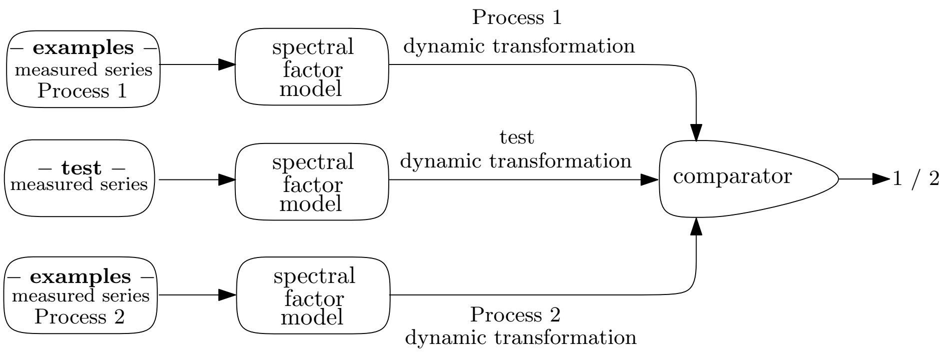

Figure 1.11: A test time series is associated to a class of time series if that class has the closest proximity, in terms of the commonalities of its examples, among all classes to the test time series.

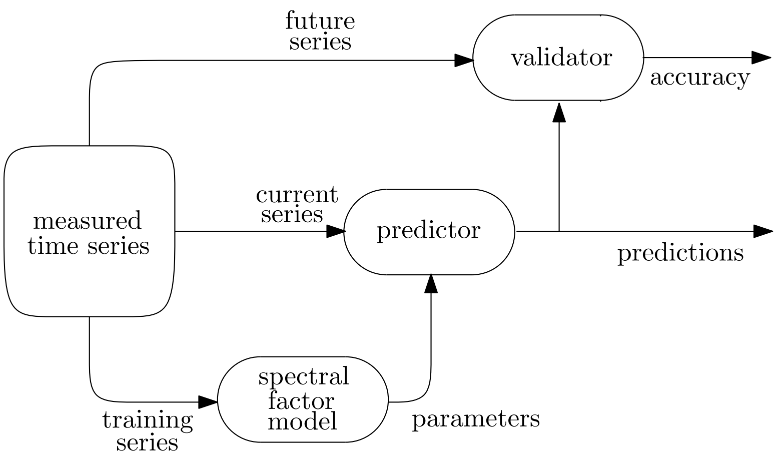

Figure 1.12: A predictor for the measured time series is built using parameters pertaining to maximally inherited commonalities of a training series. Accuracy of predictions based on current samples of the measured time series as evidence is compared with its future samples.