Chapter 7

Experiments

The notion of multivariate time series learning was introduced in Chapter 1 using two practical example problems. One was classification of human MEG signals and the other was prediction of share prices in a portfolio. It was posited that measured multivariate signals in both these problems were generated by dynamic transformation of a low-dimensional latent time series whose *acvf* characteristics are assumed known or given. For convenience, it was assumed that the latent time series is a zero-mean unit-variance white noise.
In designing a modeling framework for multivariate time series in Chapter 4, many merits and challenges in estimating the dynamic transformation in Fourier spectral domain were seen and the modeling framework was called the spectral factor model. In deriving an optimal model in Chapter 5, the following points were considered :

(a) From all possible spectral factor models, a model that is the most likely to have generated the available measured time series according to the principle of maximum likelihood was found. For the maximum likelihood spectral transformation matrix $\mathbf{W}$, through (5.5) it was found that a unique analytical solution is infeasible; whereas an iterative solution in (5.25) was obtained.

(b) From all possible maximum likelihood spectral factor models, the one which maximizes the commonalities inherited by the dynamic transformation was sought. To attain that model, a solution each for the analytical and the iterative procedures via Algorithms 3 and 4, respectively, were formulated.

Through the design of a learning framework in Chapter 6, the following were provided:

(i) A classifier, in Algorithm 7, based on κ-nearest neighbor proximity of the projection cast by the subspace defined by the optimal spectral factor model transformation of a test time series with the training examples from various classes.

(ii) A vector autoregression prediction scheme, in Algorithm 8, that replaces the *acvf* of the measured time series in the classical prediction equations with the *acvf* corresponding to the commonalities.

Each of these learning objective, viz., classification and prediction, will be experimented with in Sections 7.1 and 7.2, respectively. In both experiments, their data acquisition scheme and the general characteristics of the measured variables will be briefly explained. Importantly, limitations and advantages of these experiments with respect to

the data will be discussed.
One important aspect of a spectral factor model that was taken for granted in the theoretical development was the choice of the latent dimensionality. Hence, in the experiments, its influence on classification and prediction accuracies will be tested. Another aspect of the modeling framework that will be tested is the optimal number of subbands as required by Theorem 2.5 of the asymptotic theory of spectral estimates.

Implementation

In addition to the practicalities discussed in Section 6.1, certain implementation aspects of the experiments need to be highlighted. In conducting these experiments, the learning capabilities of the spectral factor model are to be demonstrated. To allow appropriate benchmarking, publicly available data from live fields of study will be used without much expert insights on the processes for which the data was collected. All experiments were conducted using a standard laptop with Intel Dual Core T7200 CPU (2.00GHz). Implementation of the entire estimation and the learning experiments were written using the R language [101]; the codes are intended to be made publicly available through the Comprehensive R Archive Network [4].

7.1 Classification of magnetoencephalography signals

In the first of the introductory examples in Section 1.2, the problem of dynamic factor model using the exercise of classification of wrist movements based on magnetoencephalogram (MEG) measurements was described. The task was originally part of a prestigious international competition which has concluded; its solutions have already been published and the winners were announced [1]. The typical approach of participants in the competition involved processing time series to extract certain static time and frequency domain signatures which are then fed to state-of-the-art classifiers. Nevertheless, the competition is attempted here to demonstrate the capability of the spectral factor model in utilizing much of the commonalities captured by the the latent time series presumed for the measured MEG variables for the purpose of determining the particular class of wrist movements responsible for modulating the MEG.
Briefly recap the discussion in Chapter 1 regarding what classification of time series implies: A class of time series may be regarded as an ensemble of finite length time series episodes if they are realizations of the same dynamic transformation of the same latent time series. The dynamic transformation thus represents a class of measured time series process. But remember that the dynamic transformation is such that it allows inheritance of the commonalities maximally from the measured time series. Hence, by comparing the dynamic or spectral transformation matrix of any two measured time series processes, it should be possible to decide which among them a new unclassified measured time series is closest to.

Detailed description and information of the task are available in the competition website of [1]; the data was contributed by [2]. In summary, there are $c = 4$ classes of wrist movements for which 10 MEG time series are recorded. All movements are appropriately resampled to have $\tau = 400$ samples and have similar stimulus cues and movement procedures. Independent data sets $\mathcal{D}_1$ and $\mathcal{D}_2$ are available for two human subjects; each subject produces 40 example movements per class and with 73 and 74 unlabeled test movements, respectively. The number of test movements per class per

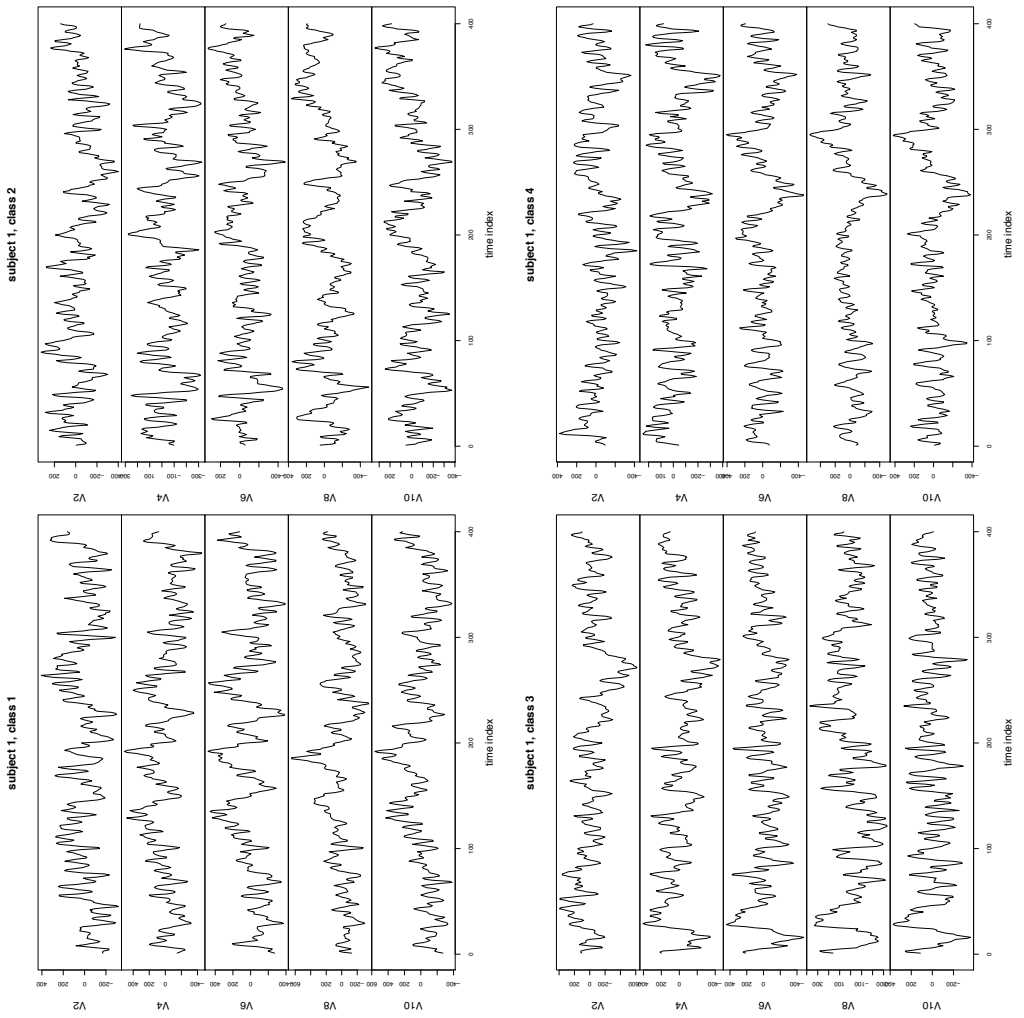

Figure 7.1: MEG signals of a human subject $\mathcal{D}_1$ corresponding to five brain spots $(V2, V4, V6, V8, V10)$ during four classes of wrist movements.

subject is also unknown. For neither learning nor testing, there is a need to mix the data from $\mathcal{D}_1$ and $\mathcal{D}_2$ whereas tests are assessed on their average count of classification accuracies, a_1 and a_2, respectively.

Testing latent dimensionality: Since it is a prerogative to estimate an appropriate latent dimensionality q for a given measured dimensionality r, the classification accuracy on all possibilities, viz., $q = 1, \ldots, r-1$ will be tested. However, as discussed in Section 6.1, there ought to be sufficient number of samples n within a subband of the discrete Fourier transform of the measured factor model for enhancing reliability of the estimated parameters $\mathbf{W} \in \mathbb{C}^{r \times q}$.

Testing number of target frequencies: Yet another constraint that was summarized in Section 6.1 was the number $\hat{j}$ of target frequencies; the sampling rate should be high enough so that sufficiently large $\hat{j}$ number of target frequencies may be assigned to meet the conditions of the asymptotic theory of spectral estimates. Unfortunately,

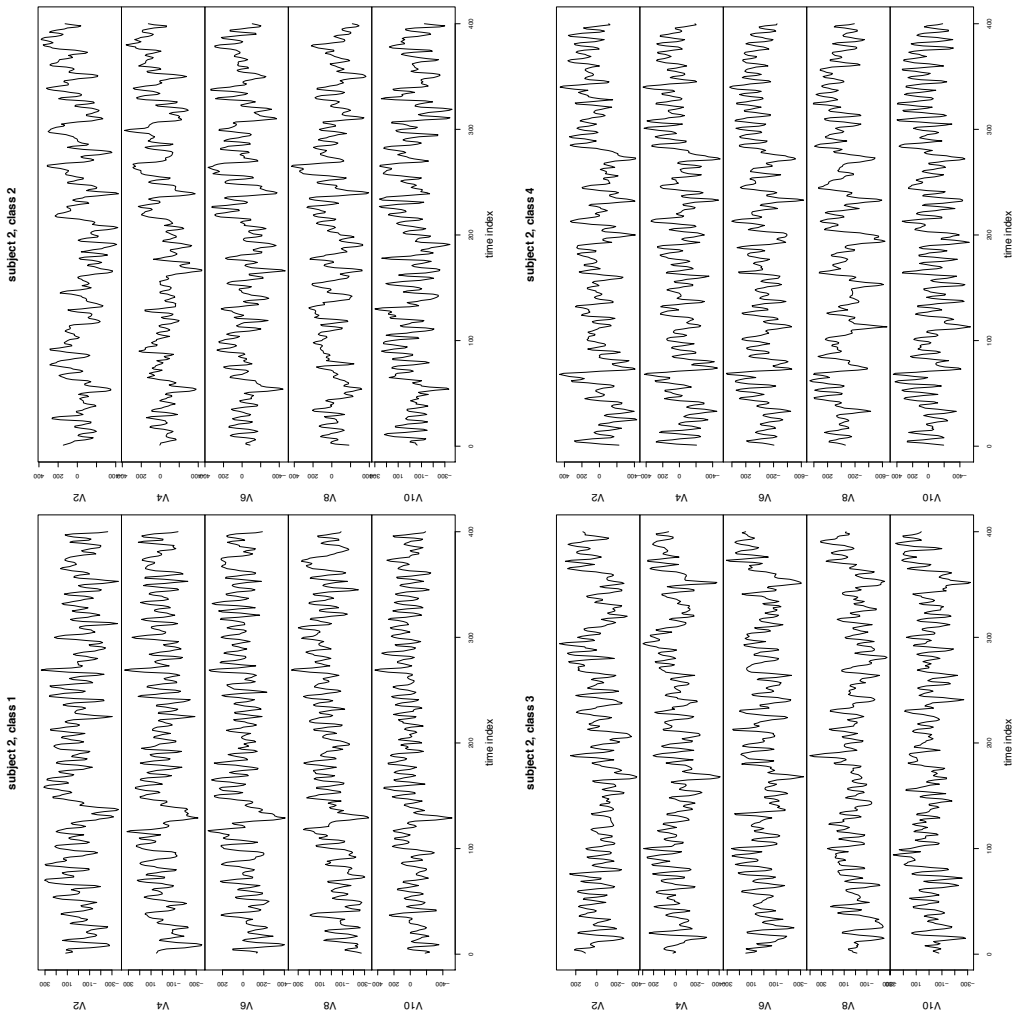

Figure 7.2: MEG signals of a human subject $\mathcal{D}_2$ corresponding to five brain spots $(V2, V4, V6, V8, V10)$ during four classes of wrist movements.

the number of samples for the data was just 400.

Balancing asymptotic Gaussanity and curse of dimensionality: A balance has to be struck between the demands for a large number of samples n within a subband for estimating the parameters for a latent dimensionality up to $q = r - 1$ while ensuring that increasing n would not hamper the large number $\hat{j}$ of target frequencies required. It is not the intention to pre-process the data to increase the sample rate or perform other modifications that might lead to explainable bias in comparison of spectral factor model performance with others. As a result, it was decided to use $r = 5$ measured signals only from among the 10 measured signals. In Figures 7.1 and 7.2, these are marked as $(V2, V4, V6, V8, V10)$ instead of $(V1, \ldots, V10)$ of Figure 1.3. The signals $V1, \ldots, V10$ correspond to spatially adjacent parts of the brain; other than that no set of signals seem qualitatively more similar to another set of signals and no particular criteria was used to select the set $(V2, V4, V6, V8, V10)$ of five

measured signals. Obviously, using only part of the measured variables for such a tedious classification exercise invites the risk of losing information rich data that might reflect in poor classification accuracy. As a validation, however, the exercise with the other set $(V1, V3, V5, V7, V9)$ of measured signals will also be carried out.

	$\hat{\jmath} = 20$		$\hat{\jmath} = 25$		$\hat{\jmath} = 30$	
q	a_1	a_2	a_1	a_2	a_1	a_2
1	40.54	32.88	40.54	32.88	40.54	26.03
2	40.54	32.88	40.54	32.88	40.54	26.03
3	40.54	32.88	40.54	32.88	39.19	31.51
4	40.54	32.88	40.54	32.88	39.19	30.14

Table 7.1: Percentages of average accuracies a_1 and a_2 in classifying $c = 4$ classes of wrist movements on two subjects $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively, based on their 5-variate MEG $(V2, V4, V6, V8, V10)$. The classifier was based on Algorithm 7 using $\kappa = 3$ for various values for the latent dimensionality q and number of target frequencies $\hat{\jmath}$.

	$\hat{\jmath} = 20$		$\hat{\jmath} = 25$		$\hat{\jmath} = 30$	
q	a_1	a_2	a_1	a_2	a_1	a_2
1	40.54	32.88	35.14	32.88	40.54	32.88
2	39.19	32.88	36.49	32.88	41.89	26.03
3	39.19	32.88	36.49	32.88	41.89	31.51
4	35.14	32.88	35.14	32.88	40.54	30.14

Table 7.2: Results of the experiments for the 5-variate MEG $(V1, V3, V5, V7, V9)$ with the same setup as in Table 7.1.

The accuracy of the classification are available in Tables 7.1 and 7.2. Note that the data obtained for both those tables are from the same set of processes with a different set of measured variables. However, within a table there are some accuracies which do not seem to change with dimensionality q or number of subbands $\hat{\jmath}$. Explaining such results is attempted below:

Class imbalance: The number of test episodes per class was unequal. Note that, had the classes were balanced, the classification is considered to be worse than random classification if accuracies a_1 and a_2 were below $\frac{1}{c} = 25\%$; whereas perfect classification will imply 100% in any case.

Nearest neighbours: Tests on $\kappa = 5$ proved to be not significantly different from those presented in Tables 7.1 and 7.2 for $\kappa = 3$. Whereas for larger κ, the accuracies were poorer especially for larger q possibly due to the sparsity of consistent training samples against a larger set of features.

Asymptotic Gaussanity The subbands tend to lose their distinct Gaussanity with increasing bandwidth, e.g., $\hat{\jmath} = 25$ and $\hat{\jmath} = 20$. In such situations, the classification accuracy becomes invariant as more Gaussian subbands are merged. Subbands could not be increased disproportionately because of the following reason.

Rank	a_1	a_2	Competing methods
1	59.5	34.3	Reported access to 'bipolar' time series unavailable to others. Fourier and wavelet features selected via genetic algorithm. Support vector and linear discriminant classifiers.
2	31.1	19.2	0-0.5 s segment with 0.5-8 Hz + 20 Hz subsampling Principal Fisher discriminant time and Fourier features. Fisher discriminant classifiers.
3	16.2	31.5	Fourier, wavelet features selected via genetic algorithm. Support vector classifiers.
4	23.0	17.8	0-0.5 s segment with 0.5-8 Hz. Principal Fisher discriminant time and Fourier features. Fisher discriminant classifiers.

Table 7.3: Percentage of average accuracies of the winners published by [1].

Curse of dimensionality: With $\tau = 400$ and 20 being the number of transformation matrix parameters for $q = 4$, the subbands with $n = 20; \hat{\jmath} = 20$; $n = 16; \hat{\jmath} = 20$; and $n \approx 13; \hat{\jmath} = 20$ will all challenge the asymptotic theory and suffer from the curse of dimensionality.

Competition: It is noteworthy that had the spectral factor model competed in [1] with any q and $\hat{\jmath}$ setting, as shown in Table 7.3, the spectral factor model would have bettered all reported accuracies except against the topper. The topper of the competition seemingly had an advantage of prior knowledge or extra information regarding the time series. Also, no pre-processing of the time series was done unlike the competitors; this is because expertise on the scientific procedure of the data acquisition was lacking nor was it desired to skew the benchmarking of the spectral factor model through unexplainable effects of data pre-processing. However, a basic Bartlett-Hann windowing [50] is performed. This is a standard procedure for discrete Fourier transform techniques to reduce the Gibbs phenomenon as the theoretically periodic finite length realization of a time series is truncated [52].

Moreover, based on available results at [1], the spectral factor model results are a clear front runner despite not requiring any of the advanced process knowledge and preprocessing of the competitors. Also, it is very likely that there was a handicap in the accuracy of the classification due to the inability to use all the measured MEG variables due to the low data sampling rate as explained earlier. Nevertheless, the results obtained demonstrate sufficient classification capabilities of the spectral factor model.

7.2 Prediction of yield rates of shares

In Section 1.2, the discussion on the setup of the prediction experiments was initiated through the example of a portfolio of shares obtained from [6]. The same motivation, data and setup are continued here. There is access to a multivariate time series consisting of synchronously sampled daily share prices of 6 German companies over a

period of 2747 trading days during 01/01/1983 - 30/12/1993. As shown in Figure 1.5, the component time series demonstrate similar dynamic covariations when they increase or decrease with observable patterns which are not necessarily readily quantifiable.

It may be verified from Figure 1.5 that there exist increasing and decreasing general trend patterns over a substantial number of samples. Hence, regression detrending on the training series [22] will be performed. The current test series subject to prediction is detrended using the parameters of the regression; the result is displayed in Figure 7.3. Despite this detrending, there still exists obvious non-stationarity in the data.

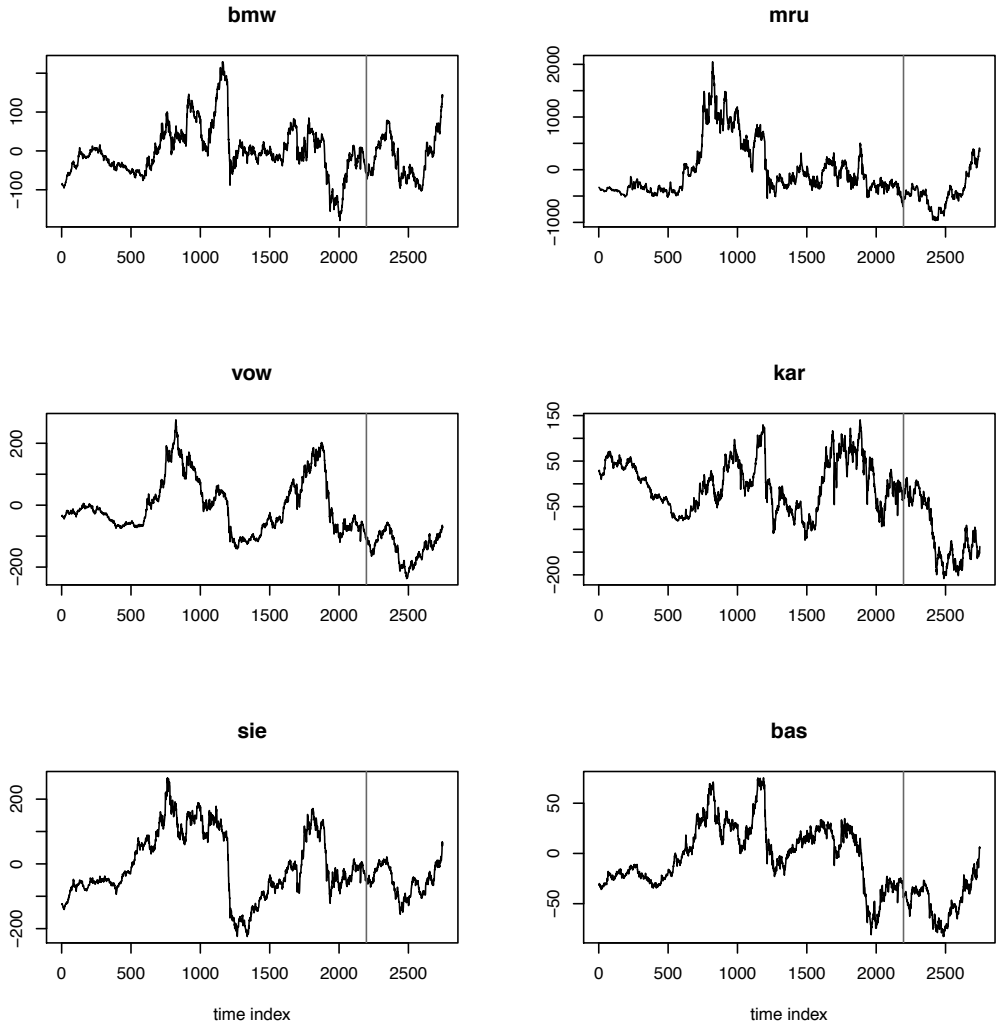

Figure 7.3: Original share prices shown in Figure 1.5 are regression detrended and split (by the gray vertical line) into the training series followed by the test series. The test series is corrected using the regression parameters of the training series.

Hence, this prediction exercise will implicitly also test the robustness of the spectral factor model in deviations from the assumption of weak stationarity.

Another pre-processing is effected in the frequency-domain for the robustness of spectral estimates. Prior to estimation of the spectral factor model, windowing of the

time series is performed to reduce the Gibbs phenomenon arising due to disparities between the ends of the finite length realization of the time series [52]. As with the previous experiment, a basic Bartlett-Hann windowing [50] will be performed to the measured time series.

How hard is this chosen problem of prediction of multivariate time series? To answer this question, the predictability of each component series of the multivariate time series has to be checked. In Algorithm 8, the estimate for the time series $\{\mathrm{y}_t\}$ for a horizon s given the current sample y_t and the past $p-1$ samples of the series was developed. On the other hand, a very naïve prediction is to assume that the future evolution is held on the current value. Obviously, the naïvety will incur errors given the stochastic nature of the time series. To measure the accuracy of the prediction, the ratio of the mean of the square errors normalized to the variance of the true time series, called the normalized mean square error (NMSE), is used. The sample counterpart of the population NMSE will be used to assess predictive performance.

naïve prediction

s	bmw	mru	vow	kar	sie	bas
1	1.24	0.83	1.33	0.96	1.53	1.8
2	2.84	1.77	2.66	1.97	3.11	3.57
3	4.65	2.78	3.94	2.97	4.8	5.25
4	6.54	3.83	5.24	3.83	6.36	6.66
5	8.58	4.84	6.55	4.72	8.14	8.14
6	10.57	5.97	7.99	5.69	10.04	9.71
7	12.47	7.07	9.35	6.73	11.81	11.12
8	14.37	8.25	10.77	7.75	13.55	12.61
9	16.36	9.48	12.34	8.86	15.40	13.91
10	18.30	10.78	13.90	9.95	17.20	15.09
20	37.09	22.56	30.79	21.08	37.44	31.08

Table 7.4: NMSE% of the naïve prediction $\hat{y}_{t+s|t} = y_t$ of each component share price of the portfolio for various horizons s.

Table 7.4 gives the NMSE for the naïve prediction of each component measured time series for various horizons s. Note that for $s = 1$, i.e., for the next trading day, the naïve prediction is reasonable as the NMSE registers just about 1% prediction error of the variance of their true evolution. For $s = 5$, which generally corresponds to a week-ahead prediction records individual prediction error NMSE averaging between 4 - 9 %, which is neither trivial nor grossly incorrect. For $s = 10$ and $s = 20$ in Table 7.4, it may be seen that the naïve prediction deteriorates substantially for larger horizons.

The spectral factor model prediction methodology is due to Algorithm 8. Following the notations in earlier chapters, the measured dimensionality is $r = 6$ and a latent dimensionality $q < r$ for the spectral factor model is presumed. Within the sufficiency of the number of samples required for a reliable estimation of transformation matrix $\mathbf{W} \in \mathbb{C}^{r\times q}$, an optimal setting of the spectral factor model was tested in trials using a part of the time series dataset for training and another for testing.

As input to Algorithm 8, the $\hat{\jmath}$ spectral factor transformation matrices $\{\mathbf{W}(\omega_j)\}, j = 1, \ldots, \hat{\jmath}$ could be provided via either Algorithms 3 or 5. As mentioned earlier, the EM algorithm requires multiple restarts and the parameters that correspond to the maxi-

mum of the converged likelihood could be chosen for maximizing the commonalities. A numerical log-likelihood convergence difference of 10^{-8} and a maximum of 20 iterations were considered appropriate [7]. For the share price portfolio dataset, the EM algorithm typically converged in less than 10 iterations and a maximum of 20 restarts were typically found appropriate in discovering a transformation matrix that is close to to 1% of the log-likelihood of the analytical solution. For the ease of reporting and with the focus on the prediction methodology, the experimental results presented here were carried out with Algorithm 3.

The following observations were made on the prediction accuracy of Algorithm 8 using the spectral factor model as measured by the NMSE on those trials:

(i) an autoregression of order $p = 2$ performed consistently much better than other orders. Hence, $p = 2$ was chosen for the experiments and presenting the results of the tests with orders $p \neq 2$ is skipped.

(ii) increasing the number $\hat{\jmath}$ of subbands of frequencies as stipulated by the asymptotic theory enhanced the prediction accuracy significantly only with $q = 1$. Hence, $\hat{\jmath} = 60$ was picked for the experiments; it corresponds to $n = 36$ discrete frequency transform components per subband which is reasonable for the estimation of the spectral factor model parameters for $r = 6$ measured time series.

It is wished to do predictions of the share prices in terms of a number of trading days, i.e., for the next day ($s = 1$), one week ahead ($s = 5$), a fortnight ahead ($s = 10$), and a month ahead ($s = 22$). Table 7.5 gives the results of the prediction exercise using the spectral factor model as per Algorithm 8 for horizons $s = 1$, $s = 5$, and $s = 10$.

It shows that increasing the latent dimensionality q increases the prediction accuracy with $q = 1$ substantially worse than others and $q = 5$ being the best. This is a logical progression of accuracy that as you increase the latent dimensionality, the commonalities of the measured time series that the spectral transformation could inherit is larger. Hence, higher the latent dimensionality, higher the accuracy or lower the NMSE.

It is the aim to pick a suitable latent dimensionality q by trading accuracy of the prediction NMSE for the number of parameters rq. It is numerically obvious from Table 7.5 that there is a significant advantage in terms of the NMSE in picking $q \notin \{1, 2\}$ but $q \in \{3, 4, 5\}$. Moreover, picking $q > 3$ seems not to improve the accuracy much. On comparing the NMSE from Table 7.4 for various horizon with Table 7.5, it is evident that the spectral factor model for $q \in \{3, 4, 5\}$ is a much more accurate long-term predictor than sample *acvf*-based classical autoregressive predictor.

Algorithm 8 recommended replacing the *acvf* Γ_h^{y} of the measured time series $\{\mathsf{y}_t\}$ with the *acvf* Γ_h^{v} of the dynamically transformed latent variables obtained through the spectral factor model estimation. As a result, spectral factor model predictions are assessed with the accuracy of the original predictions with the sample *acvf* using the classical vector autoregression of (6.10). Table 7.6 gives the NMSE% of $\hat{y}_{t+1|t}$ according to (6.10) for various orders p of autoregression. The sharp decline in the prediction of most of the component time series with increasing orders shows that the sample *acvf* estimates are very unreliable; the predictions $\hat{y}_{t+5|t}$ in Table 7.6 also corroborate such a conclusion. Moreover, on comparing Table 7.6 with Table 7.5, it is seen that for $s = 1$ the performance of the spectral factor model with $q \in \{3, 4, 5\}$ is similar in performance to the classical vector autoregression with $p = 1$. On the

spectral factor model-based vector autoregression

q	bmw	mru	vow	kar	sie	bas
$s = 1$						
1	97.79	70.25	73.77	58.32	60.39	12.65
2	5.66	9.78	6.05	6.36	3.12	5.68
3	3.47	3.99	1.60	3.69	1.52	2.36
4	2.45	2.14	1.20	3.41	1.19	1.69
5	2.36	2.05	1.17	3.32	1.17	1.64
$s = 5$						
1	39.75	94.31	205.65	15.35	174.06	15.96
2	29.11	28.55	22.78	26.06	11.10	19.52
3	9.53	9.57	4.26	9.97	4.06	5.40
4	7.36	5.56	3.24	9.42	3.36	4.19
5	7.28	5.44	3.23	9.58	3.36	4.19
$s = 10$						
1	92.17	236.98	143.46	168.84	195.85	29.74
2	52.52	30.86	9.70	63.44	11.35	10.91
3	13.96	10.27	7.05	17.02	7.83	7.95
4	13.96	9.61	6.62	16.81	6.98	7.90
5	13.51	9.43	6.36	16.73	6.74	7.72
$s = 22$						
1	191.34	390.07	213.13	310.42	458.14	96.42
2	89.85	82.62	23.09	112.64	32.94	25.56
3	37.23	24.44	17.46	33.91	23.74	18.37
4	33.37	20.90	15.11	32.29	17.68	18.15
5	32.51	20.47	14.62	33.44	17.02	17.54

Table 7.5: NMSE% of the predictions for the next day ($s = 1$), one week ahead ($s = 5$), a fortnight ahead ($s = 10$), and a month ahead ($s = 22$) for each component share price of the portfolio for various latent dimensions q; $\hat{j} = 60$ and $p = 2$ were chosen.

other hand, for $s = 5$, spectral factor model with $q \in \{3, 4, 5\}$ is clearly outperforming the classical vector autoregression. For even higher horizons of $s = 10$ and $s = 20$, the classical vector autoregression is immensely worse in performance than the spectral factor model and the results are, hence, not presented.

classical vector autoregression

p	bmw	mru	vow	kar	sie	bas
$s = 1$						
1	1.26	0.88	2.05	1.79	1.83	2.82
2	5.26	0.99	18.32	10.58	23.43	87.87
3	8.83	1.03	35.63	18.58	41.21	183.35
4	12.18	1.10	59.40	30.03	60.94	287.96
5	17.71	1.24	90.14	44.09	85.69	437.76
$s = 5$						
1	9.60	5.75	28.25	30.88	17.35	46.99
2	17.38	5.77	60.54	46.12	50.35	191.55
3	23.60	5.78	91.18	58.54	75.25	323.58
5	38.84	5.89	175.62	89.47	136.50	668.26
10	111.52	7.15	498.85	161.82	330.19	2438.70

Table 7.6: NMSE% of one day ahead ($s = 1$) and one week ahead ($s = 5$) predictions of each component share price of the portfolio for various orders p of autoregression.