Chapter 6

Learning via spectral factor model

In Chapter 1, the objective of learning a time series process was discussed with examples. Two challenges to prove the learning worth of the spectral factor model were proposed there. Firstly, a given measured time series has to be classified as belonging to one of the several possible processes that could have generated it. In this chapter, classification is done based on the proximity of the optimal spectral factor model parameters of the unclassified time series with that of the time series of various classes of possible processes. Secondly, prediction of the future evolution of a current measured time series is done. In this chapter, prediction is performed by enriching classical vector autoregression parameters of the measured time series in the prediction equation with commonalities.

In Section 6.1, before moving to either of those learning applications, it is necessary to consider the computational requirements of the spectral factor model estimation. In particular, strategies to choose the best of the two possible estimation procedures developed in Chapter 5 are considered from a practical perspective of using them in a learning problem.

In Section 6.2, the classification problem is defined concretely. The strategy involves comparing projection of the subspace spanned by the transformation matrix of the test time series episodes onto those of a number of training time series episodes. An approach based on the nearest neighbors in terms of the projection is used to decide whether a test episode belongs to one class or another; this is made available in Algorithm 7.

In Section 6.3, the prediction problem is taken up. The strategy there is simple: The measured *acvf* is an addition of two *acvf*s, one of them inheriting the commonalities and the other not. All occurrences of the measured *acvf* in the classical vector autoregression prediction equations are replaced with the part of the measured *acvf* that inherits the commonalities. This is demonstrated in Algorithm 8.

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

- ▷ Spectral factor model based classification.
  The classification metric of (6.4) compares the maximum commonality transformations of any two multivariate time series. The metric quantifies the overlap of maximum commonality subspaces despite (i) multiplicity of maximum likeli-

hood solution due to orthogonal rotations, (ii) the transformation matrices being complex-valued, and (iii) the transformations for all the subbands are to be compared. The closest work in the literature to this is that of [66] who struggle to achieve a proper metric that will compare two classes of spectral densities despite working with their full-rank sample estimates.

- ▷ Commonalities driven multivariate time series prediction.
  For predicting measured multivariate time series believed to consist of substantial commonalities, an estimate of the *acvf* is obtained by inverting its commonalities enriched spectral density function. Classical vector autoregression on current and past samples with orthogonal errors, as prevalent in literature [102, 51], is used to obtain the predictions. Except, here, the measured *acvf* is replaced by that of the commonalities estimate.
  On the other hand, the focus of eminent works in dynamic factor model literature such as [36] is in the prediction of the commonalities, which is typically unmeasured. E.g., [104] wants to model business cycles whereas [118] predicts diffusion index based on other measurable indicators.

6.1 Practicalities of spectral factor model estimation

It is clear that learning problems would require estimation of the spectral factor model parameters that inherit commonalities maximally. Hence, as a prelude to using commonalities for learning problems, Algorithm 6 is followed to estimate these parameters given a finite τ length time series $\{y_t\}$, $t = 1, 2, \ldots, \tau$. The output of the algorithm is the set of spectral model parameters $\{\mathbf{W}(\omega_j), \mathsf{S}^{\mathsf{z}}(\omega_j)\}$ at $\hat{\jmath}$ target frequencies $\omega_j \in [0, 1)$, $j = 1, \ldots, \hat{\jmath}$.

Algorithm 6: Estimate optimal spectral factor model per subband

Input: $\{y_t\}, t = 1, \ldots, \tau$; $y_t \in \mathbb{R}^r$;
Output: $\{\mathbf{W}(\omega_j)\}$; $j = 1, \ldots, \hat{\jmath}$
compute $\{\mathbf{y}(\omega_{j,l})\}$; $j = 1, \ldots, \hat{\jmath}$; $l = 1, \ldots, n$ using Algorithm 1;
foreach $j = 1, \ldots, \hat{\jmath}$ **do**
  gather $\mathcal{D} = \{\mathbf{y}(\omega_{j,l})\}$, $l = 1, \ldots, n$;
  estimate $\{\mathbf{W}(\omega_j), \mathsf{S}^{\mathsf{z}}(\omega_j)\}$ with input $\mathcal{D}$ to Algorithms 3 or 5;
end

The following observations regarding Algorithm 6 may be noted:

1. The procedures of Chapter 5 estimated the maximum commonalities spectral factor model within a spectral subband as per the asymptotic theory discussed in Section 2.5. Hence, the discrete Fourier transform components are split into $\hat{\jmath}$ subbands using Algorithm 1.

2. Each subband should have a sufficiently large n number of samples for a reliable estimation of the spectral factor model parameters; this may typically be set to $n \approx r^2$ to ensure consistency of sample estimates without inviting the curse of dimensionality issues [106, 12].

3. Although informative, the parameter S^{z} is needed for neither the classification nor the prediction exercises because it contains no commonalities which are all available through $\mathbf{W}$.

4. Depending on the computational demands and application, either Algorithms 3 or 5 may be chosen for computing the optimal parameters of the spectral factor model.

The last of the above requires further discussion. Theoretically, the analytical solution of Algorithm 3 is elegant and unique till orthogonal rotations of the transformation matrix. However, in favor of the iterative Algorithm 5 are the following practical aspects:

- ▷ For an r-variate measured time series, computing its spectral density function as well as computing its eigenvalue decomposition are typically $O(r^3)$ operations [32]. This makes Algorithm 3 very prohibitive as the number r of measured variables grows. For q-variate latent time series, the intensive operations of the EM algorithm-based estimation in Algorithm 5 are $q \times q$ matrix inverses; they are typically $O(q^3)$ operations and $q \ll r$ is the practical choice. Note that $(\mathsf{S}^{\mathsf{z}})^{-1}$ in the EM Algorithm would involve only scalar reciprocals of its diagonal. Hence, practically, for online or real-time implementations where complexity is always a constraint, spectral factor model updates could be done better using Algorithm 5.
  On the down side, as mentioned in Note 5.4, the issue of local minima in the EM algorithm poses some risk. Hence, it is desirable to confirm iterative estimates with an occasional update via Algorithm 3. Or, the randomization of the parameters in the beginning of the EM algorithm might be replaced by analytical estimates.

- ▷ In many time series, especially in econometrics, seasonality leads to distinct spikes in the spectral components. Their adjusting or correction leads to undesired consequences including elimination of true and introduction of misleading non-seasonal characteristics as well as distortion of commonalities [91]. Suppose the discrete Fourier transform components of the unadjusted seasonal time series corresponding to the suspected seasonalities are assumed missing. EM algorithm could be extended to impute the missing values using approaches such as Monte Carlo EM [123] and Stochastic Approximation EM [31]. This allows the possibility to model and learn the commonalities without inviting unnecessary pre-processing.

6.2 Multivariate time series classification

Let an r-variate measured time series be denoted by $\{\mathsf{y}_t\}$. The objective of the classification problem is to assign $\{\mathsf{y}_t\}$ to one and only one of the c exhaustive classes of time series $\mathcal{C}_i$, $i = 1, \ldots, c$. It is necessary to clarify what a class of time series means. A **class** of time series is a stochastic process, which is distinct from other processes according to an expert who has measured the time series. Such a distinction might be due to some dynamic characteristics of the time series the class is associated with that

is objectively or subjectively obvious to the expert. Or, the expert might believe that the physical process that generated a class of time series is dissimilar to others.

To ease the discussion on classification of time series, revisit the first of the two examples in Section 1.2. There, the computer gamer has to make joystick movements which require her to position the cursor from the center of the screen to any one of the four corners. During the game, the magnetoencephalography sequences corresponding to ten spatial spots in the brain were recorded via a magnetoencephalography scanner. Existence of a set of two latent signals, viz., her cognition and reaction sequences, of known general characteristics which generate the measured time series is presumed. When a joystick is moved, these latent signals must undergo a dynamic transformation corresponding to that particular class of joystick movements. In this example, an expert might have witnessed several episodes of the gamer making these four movements and understood the dynamic characteristics of the measured time series. Each episode is a finite length multivariate time series realization. Suppose access is available to a historical database of many such episodes which have been classified by the expert; they may be called the **training episodes**. It is wished to classify more episodes without the aid of the expert one by one; each of them will be called a **test episode**.

The challenge to reliably classify a test episode of a multivariate time series process based on the dynamic characteristics of a given dataset of classified training episodes is the time series **classification problem**. Hence, the classification process will have two phases: In the training phase, the summary of the dynamic characteristics of many training episodes are extracted. Obviously, the summary here implies the parameters of the spectral factor model. In the testing phase, the test episode is fed as input to the classification system. Its dynamic characteristics are compared with the dynamic characteristics of all the classes, and the most appropriate class label is given as the output of the system. Effectively, the spectral factor model parameters of the test episode is compared with those of the training episodes.

Such a classification system is indeed a learning system because of two reasons: First, the essential dynamic characteristics from all training episodes have to be appropriately summarized, which in this thesis's context will be in model parameters. Second, the classification system demonstrates the ability to use past experiences of training episodes to respond to a new test episode which it has not witnessed earlier.

Proposal for a classification system

The motivation so far has been that, firstly, each of the components of a multivariate measured time series contribute towards the commonalities shared amongst them, and, secondly, the dynamic transformation should maximally inherit the commonalities. Then, the following steps are devised in the time series classification strategy:

1. Estimate the optimal dynamic transformation for the test episode and all training episodes whereby the latent dimensionality maximally inherits the commonalities. From this thesis's perspective, this is equivalent to estimating the maximum commonalities spectral factor model as done by Algorithm 6.

2. Create a neighborhood for the test episode by computing its proximity with each of the training episodes with respect to their optimal dynamic transformation parameters.
   For this step, the proximity of the test episode has to be compared to the training

episodes with respect to their maximal inheritance of commonalities, which are believed to distinguish one class from another. The spectral factor model parameters $\mathbf{W}$ are known to correspond to maximal inheritance of commonalities. Hence, the proximity of the optimal $\mathbf{W}$ of the test episode ought to be compared with the optimal $\mathbf{W}$ of the training episodes.

3. Classify test episode to the class in which majority of the training episodes in the former's immediate neighborhood belong to.
   For this step, one should be able to design a distance metric between the transformation matrices of any two time series. Then, with respect to the distance metric, the concepts of 'neighborhood' of the transformation matrices as well as the 'closeness' between them may be used. Specifically, one may decide in favor of the class which has κ training episodes closer to the test episode than any other class; this strategy is generally known as the κ-nearest neighbor classification [28].

The last two steps above beg elaboration. Suppose access is available to time series from c classes $\mathcal{C}_i, i = 1, \ldots, c$ each with $|\mathcal{C}_i|$ examples and an unclassified time series episode. To proceed further, let the following notations be compiled:

$\{y_t\}_{l@i},\ l = 1, \ldots, \lvert\mathcal{C}_i\rvert$	l-th example time series in the class $\mathcal{C}_i$
$\{\mathbf{W}(\omega_j)\}_{l@i}$	spectral transformation of $\{y_t\}_{l@i}$ at ω_j
$\{y_t\}_?$	unclassified test episode
$\{\mathbf{W}(\omega_j)\}_?$	spectral transformation of $\{y_t\}_?$ at ω_j
$\delta\left(\{\mathbf{W}(\omega_j)\}_{l@i}, \{\mathbf{W}(\omega_j)\}_?\right)$	similarity between $\{\mathbf{W}(\omega_j)\}_{l@i}$ and $\{\mathbf{W}(\omega_j)\}_?$
$\rho(l@i, ?)$	proximity between $\{y_t\}_?$ and $\{y_t\}_{l@i}$
$\rho(\downarrow l@i, ?)$	decreasing sort of $\rho(l@i, ?)$ over l

Since the spectral factor models at $\hat{\jmath}$ target frequencies are independent of one another, it is proposed that

$$\rho(l@i, ?) = \sum_{j=1}^{\hat{\jmath}} \delta\left(\{\mathbf{W}(\omega_j)\}_{l@i}, \{\mathbf{W}(\omega_j)\}_?\right). \tag{6.1}$$

Then, $\{y_t\}_?$ may be associated with $\mathcal{C}_{\hat{\imath}}$, if

$$\hat{\imath} = \underset{i}{\text{argmax}} \sum_{l=1}^{\kappa} \rho(\downarrow l@i, ?), \tag{6.2}$$

where a tie is broken at random and κ is a suitable integer, e.g., $\kappa = 5$.

Classification metric

Recall from solution (5.10) that columns of $\mathbf{W} \in \mathbb{C}^{r \times q}$ form a set of scaled unitary vectors which define a q-dimensional space embedded in $\mathbb{C}^r$. These vectors carve a hyperparallelopiped in $\mathbb{C}^r$ whose sides are norms of these columns [112]. Then, a possible measure of disparity or similarity between any two transformation matrices is to compare the overlap of the volumes.

What the overlap of volumes really implies is specified now. The overlap for $a, b \in \mathbb{C}^r$ is defined as $\delta(a,b) = |a^*\, b|$, which is the absolute 2-norm of the unitary projection of a onto the span of b. Consider a set of linearly independent columns vectors of some matrix A spanning a subspace $\mathcal{M}_A \subset \mathbb{C}^r; \mathsf{rank}(\mathcal{M}_A) = q$ which is unitarily projected onto a subspace $\mathcal{M}_B \subset \mathbb{C}^r; \mathsf{rank}(\mathcal{M}_B) = q$ of another matrix B. This projection may be thought of as carving a volume measured as the absolute determinant $|\det(A^*B)|$ of the unitary projection of the vectors spanning $\mathcal{M}_A$ onto $\mathcal{M}_B$ [75, 84]. In [85], it is available that

$$|\det(A^*B)| = \mathsf{vol}(A)\mathsf{vol}(B)\cos\{\mathsf{R}(A), \mathsf{R}(B)\} \tag{6.3}$$

where $\mathsf{vol}(A) \triangleq \det(\mathsf{R}(A))$, where $\mathsf{R}(A)$ is the range space of A and $\cos\{\mathsf{R}(A), \mathsf{R}(B)\}$ refers to the product of the principal angles between compatible matrices A and B. In [40], it is shown that $\cos\{\mathsf{R}(A), \mathsf{R}(B)\} = \prod_{k=1}^{q} |a_k^*\, b_k|$, where a_k and b_k correspond to the k-th principal singular vector pair of A and B, respectively. For the purposes here, it is appropriate to use (6.3) to find

$$\delta\left(\{\mathbf{W}(\omega_j)\}_{l@i}, \{\mathbf{W}(\omega_j)\}_{?}\right) \triangleq |\det(\{\mathbf{W}(\omega_j)\}_{l@i}^*\, \{\mathbf{W}(\omega_j)\}_{?})|. \tag{6.4}$$

Salient features of the classification metric: The metric due to (6.1) and (6.4) is superior to those proposed by [66] for multivariate time series classification because (i) it evaluates the latent structure (ii) is invariant to orthogonal rotations of the transformation matrix, (iii) applicable in rank-deficient spectral density functions, and (iv) scalable with the number of subbands.

Classification algorithm

It is now ready to classify a test series $\{y_t\}_?$ based on class affiliations and distances to the κ-nearest neighbor training series from classes $\mathcal{C}_i, i = 1, \ldots, c$ each with $|\mathcal{C}_i|$ training series whose l-th example is $\{y_t\}_{l@i}$, $l = 1, \ldots, |\mathcal{C}_i|$. The classification procedure is simple and is given in Algorithm 7:

Algorithm 7: Spectral factor model classification

Input: $\{y_t\}_?$, $\{y_t\}_{l@i}, i = 1, \ldots, c; l = 1, \ldots, |\mathcal{C}_i|$;
Output: $\hat{\imath} : \{y_t\}_? \in \mathcal{C}_{\hat{\imath}}$
choose Algorithm 3 in Algorithm 6 and for
$j = 1, \ldots, \hat{\jmath}$
    estimate output $\{\mathbf{W}(\omega_j)\}_?$ with input $\{y_t\}_?$;
    estimate output $\{\mathbf{W}(\omega_j)\}_{l@i}$ with input $\{y_t\}_{l@i}$;
compute $\rho(l@i, ?)$ using (6.1) and (6.4);
compute $\hat{\imath}$ using (6.2);

Note 6.1. *Algorithm 3 was insisted in Algorithm 7 because the solution based on EM algorithm of Algorithm 5 does not guarantee orthogonal columns for the spectral transformation matrix* $\mathbf{W}$ *for the metric (6.4) to be directly applicable.*

Note 6.2. *The optimal parameters of the training episodes* $\{\mathbf{W}(\omega_j)\}_{l@i}$, $j = 1, \ldots, \hat{\jmath}$, $l = 1, \ldots, |\mathcal{C}_i|$ *may be computed offline and only once.*

6.3 Multivariate time series prediction

The **prediction problem**, as introduced in Chapter 1, meant reliable estimation of the future evolution of a given time series realization. Subsequently, through the spectral factor model, a parametric time series model was developed; it assumes existence of latent time series that could be dynamically transformed to imitate a higher dimensional multivariate measured time series by inheriting the commonalities of the measured variables. As a solution, it is hoped to drive the future evolution of a given realization by using the commonalities and avoiding the idiosyncarsies.

Prediction methodology

Insofar as to validate the robustness of the spectral factor model and its underlying assumptions, the intention is to predict the evolution of the time series using the commonalities the spectral factor model could extract from the data. In order to validate that the predictions are benchmarked appropriately, it is necessary to compare the prediction accuracy of the spectral factor model with those of the state-of-the-art multivariate time series models. Then, it seems reasonable to modify the parameters of the state-of-the-art models to be dependent on the commonalities only and assess the accuracy upon that modification.

Fortunately, the aforementioned modification of the state-of-the-art model in the context of this thesis is easy. This is because the spectral factor model was built on the spectral density function or equivalently on the *acvf* of stationary processes; whereas the *acvf*s decompose into parts which are commonalities-dependent and commonalities-independent as per (4.3).

Predicting a multivariate measured time series using commonalities dependent state-of-the-art prediction models accurately should strongly hint that the evolution of the time series is driven by the commonalities. Then, the assumption regarding a latent time series will stand vindicated. On the contrary, if the component time series are all uncorrelated there will not be much to gain in prediction through this approach.

Classical vector autoregressive prediction

One of the most widely used family of equivalent time series models based on classical **vector autoregressive modeling** of linear processes will be used [102, 51]. This is because the prediction framework in that model is simple to comprehend, popularly tested, and easy to implement. Later, the classical model will be adapted such that its parameters are maximal carriers of commonalities.

The basic principle of vector autoregression is to estimate a future sample of a given realization as a weighted sum of the current and past samples. One may refer [51] among many references to pick from a wide ranging approaches ranging from maximum likehihood estimation, Kalman filter, Bayesian analysis, etc. to time series prediction; in moving forward, just one of those approaches based on linear projections is used. For now, the classical vector autoregressive model may be summarized as follows: For an r-variate linear process $\{\mathrm{y}_t\}$ up to the current sample y_t, a simple and valuable version of the prediction problem involves estimating the s-th next sample

$y_{t+s|t}$ as a linear function of a finite number p of the present and past samples as

$$\mathsf{y}_{t+s|t} = \epsilon_{t+s} + \sum_{j=0}^{p-1} \phi^{\mathsf{y}}_{j+1,s} \mathsf{y}_{t-j}, \tag{6.5}$$

where $\epsilon_{t+i} = \mathsf{y}_{t+i} - \mathsf{y}_{t+i|t}\, \forall i = 1, 2, \ldots$ is the estimation error and $\phi^{\mathsf{y}}_{l,s} \in \mathbb{R}^{r\times r}\, \forall l = 1, 2, \ldots, p$ are the autoregression coefficient matrices. The condition that ensures minimum mean square error are when the errors ϵ_{t+i} above are uncorrelated, i.e., $\mathsf{E}[\epsilon_{t+i}\mathsf{y}'_{t-j}] = 0\ \forall j = 0, \ldots, p-1$; refer, e.g., Theorem 4.5 of [48], for this well-known result. It gives rise to the relation between the *acvf*s and the coefficient matrices:

$$\Phi^{\mathsf{y}}_{p,s} = [\phi^{\mathsf{y}}_{1,s}\, \phi^{\mathsf{y}}_{2,s}\, \cdots\, \phi^{\mathsf{y}}_{p,s}]' = (\Xi^{\mathsf{y}}_p)^{-1} \rho^{\mathsf{y}}_{p,s}, \tag{6.6}$$

where

$$\rho^{\mathsf{y}}_{p,s} = \left[\Gamma^{\mathsf{y}}_s\, \Gamma^{\mathsf{y}}_{s+1} \cdots \Gamma^{\mathsf{y}}_{s+p-1}\right]', \tag{6.7}$$

and

$$\Xi^{\mathsf{y}}_p = \begin{bmatrix} \Gamma^{\mathsf{y}}_0 & \Gamma^{\mathsf{y}}_1 & \cdots & \Gamma^{\mathsf{y}}_{p-1} \\ \Gamma^{\mathsf{y}}_{-1} & \Gamma^{\mathsf{y}}_0 & \cdots & \Gamma^{\mathsf{y}}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma^{\mathsf{y}}_{-p+1} & \Gamma^{\mathsf{y}}_{-p+2} & \cdots & \Gamma^{\mathsf{y}}_0 \end{bmatrix}. \tag{6.8}$$

In practical problems of interest $(\Xi^{\mathsf{y}}_p)^{-1}$ will exist. Therefore, for any given p-length subsequence of $\{\mathsf{y}_t\}$ written as

$$\vec{y}_{t,p} = \mathsf{vec}(y_t, y_{t-1}, \ldots, y_{t-p+1}) \in \mathbb{R}^{pr}, \tag{6.9}$$

for the classical vector autoregression on past samples as per (6.5), referring to §4.3 of [51], the minimum mean square error prediction is

$$\hat{y}_{t+s|t} = \Phi^{\mathsf{y}}_{p,s}\, \vec{y}_{t,p}. \tag{6.10}$$

Spectral factor model prediction

Based on the prediction methodology envisaged in Section 6.3, Γ^{y}_h may be replaced in (6.7) and (6.8) by the part of the *acvf* which inherits the commonalities. The spectral factor model was developed based on the decomposition of the measured multivariate time series y_t as per (4.2) to v_t and z_t, which inherit the commonalities and the idiosyncrasies, respectively. The decomposition (4.3) of the *acvf* of Γ^{y}_h into Γ^{v}_h and Γ^{z}_h was also seen. It was further found out that the best approximation of Γ^{y}_h in the sense of inheriting the commonalities is Γ^{v}_h obtained via the spectral factor model. The optimal spectral factor model parameter S^{v} is related to Γ^{v}_h through (4.11).

Suppose a maximum commonalities spectral factor model is computed based on a training set of measured time series either via the analytical approach of Section 5.1 or the iterative approach of Section 5.2 according to Algorithm 6. As a result, the optimal transformation matrices $\{\mathbf{W}(\omega_j)\}, j = 1, \ldots, \hat{\jmath}$ at $\hat{\jmath}$ target frequencies may be assumed available. Then, given any subsequence $\vec{y}_{t,p}$ of the measured time series, by replacing Γ^{y}_h with Γ^{v}_h in the prediction equations, predictions may be performed as per Algorithm 8:

Algorithm 8: Spectral factor model prediction

Input: $\vec{y}_{t,p}$; $\{\mathbf{W}(\omega_j)\}, j = 1, \ldots, \hat{j}$
Output: $\hat{y}_{t+s|t}$
compute Γ_h^{v} using (4.11);
compute $\rho_{p,s}^{\mathsf{v}}$ replacing $\Gamma_h^{\mathsf{y}} \to \Gamma_h^{\mathsf{v}}$ in (6.7);
compute Ξ_p^{v} replacing $\Gamma_h^{\mathsf{y}} \to \Gamma_h^{\mathsf{v}}$ in (6.8);
compute $\Phi_{p,s}^{\mathsf{v}} = (\Xi_p^{\mathsf{v}})^{-1} \rho_{p,s}^{\mathsf{v}}$;
estimate $\hat{y}_{t+s|t} = \Phi_{p,s}^{\mathsf{v}} \, \vec{y}_{t,p}$;

6.4 Summary

In practical learning problems one is bound to use spectral factor model with limited computational resources. In Section 6.1, choosing the estimation procedure was discussed; it was based on either (i) the cheaper EM algorithm but with necessary caution to evade local optimum traps or (ii) the accurate but expensive analytical formulas of the low-rank approximation.

For classification of multivariate time series based on the similarities of their commonalities, a metric in (6.1) and a κ-nearest neighbor classification rule in (6.2) was designed. A test multivariate time series may be classified as belonging to the class of training multivariate time series for which the subspaces spanned by their optimal spectral factor transformation matrices overlap maximally.

For prediction of multivariate time series based on the spectral factor model, the classical vector autoregression prediction models was modified by replacing the measured *acvf* with the *acvf* corresponding to the optimal commonalities.