Identifying secondary series for stepwise common singular spectrum analysis

H Viljoen[*]   SJ Steel[†]

Abstract

Common singular spectrum analysis is a technique which can be used to forecast a primary time series by using the information from a secondary series. Not all secondary series, however, provide useful information. A first contribution in this paper is to point out the properties which a secondary series should have in order to improve the forecast accuracy of the primary series. The second contribution is a proposal which can be used to select a secondary series from several candidate series. Empirical studies suggest that the proposal performs well.

1   Introduction

Common singular spectrum analysis (CSSA) was introduced by Viljoen & Nel (2010). It is a multiple time series analysis procedure, combining singular value decomposition of an unfolding matrix with the concept of common principal components introduced by Flury (1988). A modification of CSSA is proposed by Viljoen & Steel (2012). It entails replacing the original maximum likelihood approach by a stepwise procedure which was introduced by Trendafilov (2010). Another popular procedure for multiple time series analysis is multi-channel singular spectrum analysis (MSSA), an extension of singular spectrum analysis (Golyandina *et al.* 2001, Golyandina & Stepanov 2005). The primary focus in this paper is on (stepwise) CSSA, but MSSA and SSA are also considered for comparison purposes in the simulation study discussed in §3.

Although CSSA caters for an arbitrary number of time series we restrict our discussion to the case of two series. In such a context it frequently makes sense to distinguish between a primary and a secondary series. The main interest is in forecasting the primary series and to improve the forecast accuracy by using the information provided by the secondary

---

[*]Corresponding author: Department of Statistics and Actuarial Sciences, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa, email: lienki@sun.ac.za

[†]Department of Statistics and Actuarial Sciences, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa.

series. An important question that arises, concerns the properties required for a secondary series to provide useful information for improving forecasting accuracy. This question is addressed in §3. A simulation study is described in which the improvement in forecast accuracy achieved by different secondary series was investigated for CSSA and MSSA. In practical applications one often has multiple candidate secondary series and a next important question is whether a strategy can be proposed for identifying a good secondary series. In §4 an answer is provided to this question. The proposed strategy is based on the residuals obtained by combining the primary series with every candidate secondary series in a pairwise CSSA. In §5 the proposal is illustrated with a practical example. Section 6 contains concluding remarks and ideas for further research.

In the next section background information on MSSA as well as stepwise CSSA is given.

2 MSSA and stepwise CSSA

Consider two time series of equal length N, y_{1t} and y_{2t}, where $t = 1, \ldots, N$. A signal-plus-noise structure is assumed for both series, *i.e.*

$$y_{jt} = f_{jt} + \varepsilon_{jt} \qquad j = 1, 2, \text{ and } t = 1, \ldots, N. \tag{1}$$

Here, f_j is the systematic (signal) component for series j, $j = 1, 2$, and ε_j is the random (noise) component. Later it is assumed that ε_1 and ε_2 are Gaussian random variables. Our primary interest is in forecasting f_1 and we intend using information supplied by the second time series to improve the accuracy of our forecasts.

Both MSSA and stepwise CSSA consider these time series in terms of trajectory matrices. Let L, called the window length or embedding dimension, be an integer satisfying $1 < L < N/2$. We form the following $L \times K$ trajectory matrices

$$\boldsymbol{Y}_j = \begin{bmatrix} y_{j1} & y_{j2} & \cdots & y_{jK} \\ y_{j2} & y_{j3} & \cdots & y_{j,K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{jL} & y_{j,L+1} & \cdots & y_{jN} \end{bmatrix}, \qquad j = 1, 2,$$

for the two series. Here, $K = N - L + 1$. Note that $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are Hankel matrices. The trajectory matrix for the multivariate time series is $\boldsymbol{Y} = [\boldsymbol{Y}_1 \ \boldsymbol{Y}_2]$.

MSSA is based on the singular value decomposition of $\boldsymbol{Y}^T$. Let $\boldsymbol{Y}^T = \boldsymbol{U}\boldsymbol{A}\boldsymbol{V}^T$ denote this decomposition. In this expression, $\boldsymbol{U}$ is a $2K \times L$ matrix, the columns of which span the column space of $\boldsymbol{Y}^T$, $\boldsymbol{V}$ is an $L \times L$ matrix, the columns of which span the row space of $\boldsymbol{Y}^T$, and $\boldsymbol{A}$ is an $L \times L$ diagonal matrix of singular values $a_1 \geq a_2 \geq \cdots \geq a_L \geq 0$. Also, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_L$. The matrix $\boldsymbol{Y}$ can be expressed in terms of the singular values and the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, *viz.* $\boldsymbol{Y} = \sum_{i=1}^{L} a_i \boldsymbol{v}_j \boldsymbol{u}_i^T$, and with d denoting the number of positive singular values, this becomes $\boldsymbol{Y} = \sum_{i=1}^{d} \boldsymbol{Y}^{(i)}$, where $\boldsymbol{Y}^{(i)} = a_i \boldsymbol{v}_j \boldsymbol{u}_i^T$, for $i = 1, \ldots, L$. The matrices $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(d)}$ are bi-orthogonal and of rank 1. MSSA proceeds by splitting the matrices $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(d)}$ into two groups, $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(r)}$ and $\boldsymbol{Y}^{(r+1)}, \ldots, \boldsymbol{Y}^{(d)}$, with the first group corresponding to estimated signal and the second group to noise or

residual. The parameter r has a strong effect on the results of MSSA as will be seen in the simulation results later.

With r given, let $\boldsymbol{F} = [\widehat{\boldsymbol{F}}_1 \;\; \widehat{\boldsymbol{F}}_2] = \sum_{i=1}^{r} \boldsymbol{Y}^{(i)}$. This matrix may be seen as a smoothed version of $\boldsymbol{Y}$, with $\widehat{\boldsymbol{F}}_1$ corresponding to $\boldsymbol{Y}_1$ and $\widehat{\boldsymbol{F}}_2$ to $\boldsymbol{Y}_2$. The matrices $\widehat{\boldsymbol{F}}_1$ and $\widehat{\boldsymbol{F}}_2$ will typically not have a Hankel structure, and we therefore require a further step, *reverse diagonal averaging*, to obtain estimates of the signals corresponding to the two time series (Golyandina *et al.* 2001, pp. 17). Applying this to each of the matrices $\widehat{\boldsymbol{F}}_1$ and $\widehat{\boldsymbol{F}}_2$, we obtain estimates $\hat{f}_{1t}$ and $\hat{f}_{2t}$ for $t = 1, \ldots, N$ of the signal series corresponding to the two observed time series.

Turning our attention to forecasting, the following approach was proposed by Venter (2000) and is equivalent to the approach proposed by Golyandina *et al.* (2001). We consider forecasting f_1 using the matrix $\widehat{\boldsymbol{F}}^{(1)} = [\hat{\boldsymbol{f}}_1^{(1)}, \ldots, \hat{\boldsymbol{f}}_K^{(1)}]$. A one-step-ahead forecast can be obtained by appending another column $\hat{\boldsymbol{f}}_{K+1}^{(1)}$ to this matrix. This column has components $\hat{f}_{K+1}^{(1)}, \hat{f}_{K+2}^{(1)}, \ldots, \hat{f}_N^{(1)}, \hat{f}_{N+1}^{(1)}$, where $\hat{\boldsymbol{f}}_{N+1}^{(1)}$ is the forecast value which has to be determined. The forecast value $\hat{f}_{N+1}^{(1)}$ is found to minimize the distance between $\hat{\boldsymbol{f}}_{N+1}^{(1)}$ and the subspace spanned by the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$. Details in this regard are given for example by Golyandina *et al.* (2001, pp. 95–96). Once $\hat{f}_{N+1}^{(1)}$ has been computed, the process can be iterated to obtain recurrent forecasts for longer horizons.

In CSSA we consider k multivariate populations with $L \times L$ symmetric, positive-definite covariance matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$. The notion of common principal components (CPCs), introduced by Flury (1988), implies that these matrices share principal components, *i.e.* there exists an orthogonal matrix $\boldsymbol{\Psi}$ such that $\boldsymbol{\Psi}^T \boldsymbol{\Sigma}_i \boldsymbol{\Psi} = \boldsymbol{\Lambda}_i^2$ for $i = 1, \ldots, k$. In this expression $\boldsymbol{\Lambda}_1^2, \ldots, \boldsymbol{\Lambda}_k^2$ are diagonal matrices containing the positive eigenvalues of $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$ respectively.

In practice, data have to be used to estimate $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$. Let $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k$ denote the usual sample covariance matrices. Estimating the CPCs entails finding an $L \times L$ orthogonal matrix $\boldsymbol{Q}$ (an estimate of the matrix $\boldsymbol{\Psi}$) and diagonal matrices $\boldsymbol{D}_1^2, \ldots, \boldsymbol{D}_k^2$ (containing the sample eigenvalues) such that $\boldsymbol{Q}^T \boldsymbol{S}_i \boldsymbol{Q} \simeq \boldsymbol{D}_i^2$ holds simultaneously for $i = 1, \ldots, k$ as closely as possible in some sense. Flury (1988) derived a maximum likelihood approach for estimating the CPCs, while Trendafilov (2010) introduced a stepwise algorithm for this purpose — the interested reader is referred to Trendafilov (2010) for more details.

The stepwise approach to CPCs can easily be used in a singular spectrum analysis framework. This entails centering the trajectory matrices $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ by subtracting the column averages, thereby obtaining $\widetilde{\boldsymbol{Y}}_1$ and $\widetilde{\boldsymbol{Y}}_2$. The symmetric matrices $\boldsymbol{S}_1 = \widetilde{\boldsymbol{Y}}_1 \widetilde{\boldsymbol{Y}}_1^T$ and $\boldsymbol{S}_2 = \widetilde{\boldsymbol{Y}}_2 \widetilde{\boldsymbol{Y}}_2^T$ are formed and can now be provided as input to the stepwise CPC algorithm. The remainder of the analysis proceeds as described in MSSA.

3   Secondary series properties

In this section the properties which render a secondary series useful for inclusion in a multiple time series forecasting scheme, such as CSSA and MSSA are determined. For

this purpose, stepwise CSSA, MSSA and SSA were applied in several simulation scenarios in order to determine when and to what extent the first two procedures improved upon SSA.

Consider two time series y_{1t} and y_{2t} for $t = 1, \ldots, N$. The signal-plus-noise structure given in (1) is assumed, with ε_1 and ε_2 distributed as follows

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right).$$

Five covariance matrices were considered in Table 1 for the noise by varying the noise variance of the secondary series and the correlation between the two noise variables.

Covariance matrix	σ_1^2	σ_2^2	ρ
1	1	1	0
2	1	0.5	0
3	1	4	0
4	1	1	0.5
5	1	1	−0.5

Table 1:   *Covariance matrix scenarios for the noise in the simulation scenarios.*

The primary signal series is assumed to be

$$f_1(t) = 2t + 5\cos\left(\frac{2\pi t}{12}\right) + 5\sin\left(\frac{2\pi t}{12}\right) \qquad\qquad t = 1, \ldots, N \qquad\qquad (2)$$

whereas the secondary signal is assumed to be

$$f_2(t) = a + b\cos\left(\frac{2\pi t}{c}\right) + b\sin\left(\frac{2\pi t}{c}\right) \qquad\qquad t = 1, \ldots, N. \qquad\qquad (3)$$

The *trend* in the secondary signal is determined by a, the *amplitude* of the signal by b and the *period* by c. Using different values for a, b and c, different secondary signals are created and then used in MSSA and stepwise CSSA to forecast the primary signal. The noise structures in Table 1 are used with each pair of primary and secondary series. This leads to the following thirty cases of secondary series as shown in Table 2.

Covariance matrix	Trend (a)	Amplitude (b)	Period (c)
1–5	$2t$	5	12
6–10	0	5	12
11–15	$2t$	5	6
16–20	$2t$	5	18
21–25	$2t$	3	12
26–30	$2t$	10	12

Table 2:   *The thirty case scenarios of secondary series.*

For Cases 1–5 the secondary series signal is identical to that of the primary series. In Cases 6–10 the secondary signal is stationary with amplitude and period equal to that of the primary series. In Cases 11–15 and 16–20 the signals of the primary and secondary

series differ with respect to period, while in Cases 21–25 and 26–30 the difference is with respect to amplitude.

In the simulation study the time series length N was set at 100 and the number of simulation repetitions was taken as 500. The six step ahead root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{6} \sum_{t=N+1}^{N+6} (f_t - \hat{f}_t)^2}. \tag{4}$$

was calculated as a measure of forecast accuracy. Note that RMSE is calculated as the square root of the average (over the six steps ahead forecast period) squared difference between the forecast values ($\hat{f}_t$ which is the same as $\hat{y}_t$) and the true signal values (f_t). This can be done since the values of f_t are known in a simulation study. The forecast errors were calculated for $L = 12$, $L = 24$ and $L = 48$. The results for $L = 12$ are omitted since there was no improvement. However, taking $L = 24$ or $L = 48$ leads to considerable improvement in the performance of all three techniques. For $L = 24$ and $L = 48$, we investigated $r = 2, 3, \ldots, 16$ and $r = 2, 3 \ldots, 25$ respectively. Table 3 contains the smallest six step ahead forecast RMS errors and corresponding values of r for each of the three techniques and the thirty cases for $L = 24$.

What conclusions can be drawn from Table 3? Overall, it is clear that CSSA is best in all cases. There are several cases where SSA outperforms MSSA, namely when the secondary series period differs from that of the primary series. MSSA is particularly bad when the secondary series period is larger than that of the primary series (see Cases 16–20). Recall that successive groups of five cases in Table 2 correspond to specific assumptions regarding the signals. For example, in Cases 1–5 the two signals are identical, in Cases 6–10 the two signals differ with respect to trend, *etc.* The five cases within each of these groups correspond to the noise covariance structures in Table 1, respectively. If we first consider the different groups, we see that the smallest forecast errors for CSSA are for Cases 1–5 and 26–30. These correspond to the scenarios where the two signals are identical (Cases 1–5) and where the amplitude of the secondary series is larger than that of the primary series (Cases 26–30). Looking in Table 3 within each of the groups, the most favourable cases for CSSA are typically the second, fourth and fifth. These correspond to a smaller secondary noise variance, a positive correlation between the two noise variables and a negative correlation between the two noise variables, respectively. It would seem that CSSA benefits especially from the information provided by the secondary series if the secondary series has the same signal structure as the primary one and/or the secondary noise variable is correlated with the primary noise variable and/or has a smaller variance. The best r values for CSSA are equal to 3 except for Cases 11–20, where $r = 5$ is best. For MSSA the optimal r value is 4 in all cases except for Cases 11–20 where $r = 6$. The best r-value for SSA is $r = 4$.

The $L = 48$ results show that the forecast errors for CSSA are largely the same whereas those of MSSA and SSA are slightly smaller. Overall, CSSA is still best in all cases. The performance of MSSA in Cases 16–20 is significantly better than with $L = 24$ and is more or less the same as that of SSA. The best scenarios for CSSA are once again Cases 1–5 and 26–30. Finally, within each group of five cases it is again the second, fourth and fifth

	Stepwise CSSA		MSSA		Stepwise SSA	
Case	r-value	RMSE	r-value	RMSE	r-value	RMSE
1	3	0.413	4	0.564	4	0.647
2	3	0.359	4	0.628	4	0.700
3	3	0.460	4	0.651	4	0.639
4	3	0.385	4	0.604	4	0.633
5	3	0.376	4	0.527	4	0.638
6	3	0.430	4	0.609	4	0.624
7	3	0.423	4	0.633	4	0.661
8	3	0.569	4	0.590	4	0.627
9	3	0.429	4	0.612	4	0.620
10	3	0.400	4	0.594	4	0.635
11	5	0.532	6	0.719	4	0.687
12	5	0.511	6	0.692	4	0.665
13	5	0.553	6	0.743	4	0.691
14	5	0.506	6	0.703	4	0.655
15	5	0.501	6	0.662	4	0.669
16	5	0.595	6	1.117	4	0.635
17	5	0.581	6	1.073	4	0.658
18	5	0.656	6	1.193	4	0.666
19	5	0.605	6	1.140	4	0.623
20	5	0.585	6	1.080	4	0.648
21	3	0.392	4	0.621	4	0.666
22	5	0.493	4	0.638	4	0.671
23	3	0.478	4	0.660	4	0.632
24	3	0.480	4	0.630	4	0.650
25	3	0.382	4	0.603	4	0.699
26	3	0.403	4	0.614	4	0.707
27	3	0.389	4	0.578	4	0.680
28	3	0.430	4	0.593	4	0.622
29	3	0.407	4	0.597	4	0.660
30	3	0.342	4	0.494	4	0.602

Table 3:   *Optimal r-values and forecast errors for $L = 24$.*

that perform best. The r-values are the same as for $L = 24$. Since the conclusions for $L = 48$ are essentially the same as for $L = 24$, we do not include the table for $L = 48$.

4   Procedures for selecting a secondary series

Consider a practical situation of forecasting a primary time series. Incorporating the information provided by a secondary series may improve the forecast accuracy of the primary series. If several secondary series are available the question arises as to which and how many of these should be included in the analysis. In this study an attempt to provide a partial answer is presented. The attention is restricted to the possibility of including a single secondary series and since CSSA was found best in the first simulation study, only this technique is used in the second simulation study. More specifically, a primary time series together with nine secondary series which are candidates for inclusion in the analysis are simulated. Based on the results of the first simulation study four procedures for selecting a secondary series is considered. For explanatory purposes the primary series

is referred to as Series 1 and to the nine secondary series as Series 2 up to Series 10. The first step is to perform CSSA with Series 1 and each of Series 2 up to Series 10. This gives nine sets of pairs of residuals (residuals for Series 1 paired with each of the Series 2–10). Selecting a secondary series makes use of these residuals. In Procedure 1 three (out of nine) secondary series with smallest residual variances is identified. Procedure 1 then selects, from these three, the one with residuals having the largest negative correlation with the primary residuals. Procedure 2 selects, from among the nine available, the secondary series with smallest residual variance. Similarly Procedure 3 selects the secondary series with residuals having the largest negative correlation with the primary residuals. Finally, Procedure 4 randomly selects a secondary series from the nine options. This procedure is included to see whether using a criterion to select the secondary series does better than simply randomly adding one. SSA is also included in order to verify the value of using a secondary series.

Seven cases were investigated in the simulation study. In all of these the primary series signal is given by equation (2). The secondary series signal is given by equation (3). Table 4 summarizes the choices for a, b and c.

Cases	a	b	c
1	$-2t$	5	12
2	0	5	12
3	$-2t$	5	12
4	$-2t$	5	6
5	$-2t$	5	18
6	$-2t$	3	12
7	$-2t$	10	12

Table 4: *Choices for a, b and c in the second simulation study.*

In each of these cases observations of the primary and nine candidate secondary series was simulated by adding Gaussian noise to the signal values. The noise values were simulated from a ten variable Gaussian distribution with mean zero and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.35 & 0.247 & 1 & -0.35 & -0.247 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.35 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.247 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ -0.35 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -0.247 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

Regarding the variances of the noise of the candidate series, we look at cases having the same noise variance as the primary series, having smaller noise variance and having larger noise variance. As far as the correlation between the primary noise series and the secondary noise series is concerned, we have three uncorrelated cases, three cases with positive correlation and three cases with negative correlation. The specific numerical

values in the matrix were used to ensure that the covariance matrix remains positive definite. The results discussed below are for $L = 24$, $N = 100$ and using 800 simulation repetitions. Root mean square errors (RMSE) are given for the four procedures and seven cases in Table 5.

		Value of r								
		2	3	4	5	6	7	8	9	10
Case 1	Proc 1	3.105	0.363	0.419	0.497	0.558	0.614	0.675	0.749	0.809
	Proc 2	3.112	0.372	0.419	0.505	0.562	0.620	0.679	0.738	0.795
	Proc 3	3.112	0.373	0.421	0.511	0.571	0.632	0.686	0.746	0.811
	Proc 4	3.138	0.409	0.446	0.524	0.570	0.642	0.698	0.767	0.825
	SSA	10.882	6.741	0.664	0.723	0.794	0.873	0.956	1.028	1.084
Case 2	Proc 1	33.732	0.415	0.467	0.547	0.605	0.664	0.727	0.789	0.844
	Proc 2	36.820	0.486	0.526	0.600	0.657	0.712	0.767	0.805	0.869
	Proc 3	34.601	0.456	0.496	0.569	0.612	0.673	0.736	0.790	0.845
	Proc 4	18.348	0.473	0.509	0.593	0.639	0.709	0.758	0.805	0.864
	SSA	10.85	6.715	0.646	0.697	0.779	0.849	0.919	1.001	1.066
Case 3	Proc 1	3.025	0.385	0.438	0.516	0.568	0.637	0.691	0.749	0.810
	Proc 2	3.032	0.424	0.471	0.547	0.600	0.670	0.727	0.790	0.860
	Proc 3	3.037	0.404	0.458	0.542	0.601	0.664	0.715	0.775	0.831
	Proc 4	3.049	0.405	0.446	0.532	0.586	0.653	0.714	0.786	0.856
	SSA	10.802	6.657	0.670	0.744	0.817	0.886	0.960	1.043	1.116
Case 4	Proc 1	5.193	4.315	2.839	0.493	0.535	0.607	0.658	0.727	0.782
	Proc 2	5.052	3.834	2.652	0.512	0.551	0.620	0.670	0.732	0.794
	Proc 3	5.165	4.207	2.834	0.512	0.551	0.620	0.671	0.735	0.785
	Proc 4	4.680	3.909	2.314	0.521	0.557	0.616	0.674	0.749	0.809
	SSA	10.815	6.686	0.638	0.701	0.770	0.842	0.910	0.978	1.042
Case 5	Proc 1	6.659	6.518	3.090	0.572	0.608	0.688	0.733	0.798	0.850
	Proc 2	6.658	6.497	3.086	0.583	0.612	0.684	0.738	0.797	0.858
	Proc 3	6.652	6.506	3.090	0.593	0.623	0.691	0.744	0.813	0.872
	Proc 4	6.662	6.513	3.098	0.604	0.639	0.708	0.747	0.825	0.891
	SSA	10.846	6.716	0.664	0.724	0.811	0.879	0.943	1.011	1.089
Case 6	Proc 1	3.102	0.371	0.416	0.500	0.56	0.625	0.678	0.744	0.797
	Proc 2	3.126	0.401	0.449	0.520	0.569	0.634	0.687	0.768	0.823
	Proc 3	3.124	0.396	0.437	0.511	0.571	0.645	0.697	0.757	0.812
	Proc 4	3.119	0.430	0.467	0.539	0.591	0.658	0.713	0.779	0.828
	SSA	10.820	6.695	0.649	0.701	0.787	0.854	0.932	1.006	1.071
Case 7	Proc 1	3.112	0.363	0.417	0.499	0.566	0.630	0.694	0.764	0.820
	Proc 2	3.116	0.375	0.430	0.508	0.571	0.632	0.695	0.758	0.815
	Proc 3	3.099	0.381	0.431	0.506	0.558	0.633	0.696	0.769	0.842
	Proc 4	3.142	0.397	0.442	0.524	0.579	0.660	0.708	0.783	0.829
	SSA	10.835	6.701	0.681	0.739	0.814	0.887	0.968	1.042	1.116

Table 5: *Forecast RMSE for different values of r.*

Figures 1 and 2 illustrate the series that were selected by the different procedures in the seven different cases. Given the best r-value, Procedure 1 almost always selects Series 8. This is not unexpected since Procedure 1 is designed to select the secondary series simultaneously having small residual variance and with residuals having negative correlation with the primary series residuals. Series 8 is simulated to have both these properties. As expected the residuals from the pairwise CSSA analyses to a large extent reflect the covariance structure used in generating the residuals of the series (provided that the correct value of r is chosen). Similar remarks can be made for Procedures 2 and 3: Procedure

2 chooses either series 2, 5 or 8 (corresponding to medium residual variances equal to 1) while Procedure 3 chooses series 9 (the series with residuals having the largest negative correlation with the primary series residuals) with high probability. As expected, Procedure 4 almost uniformly selects any of the series.

In Figure 3 the choice of secondary series for Procedure 1 and Case 4 is shown for different values of r. We see that initially there is no clear pattern. From $r = 4$ onwards there is a very strong preference for Series 8. Note that $r = 5$ is the best choice of r for this case.
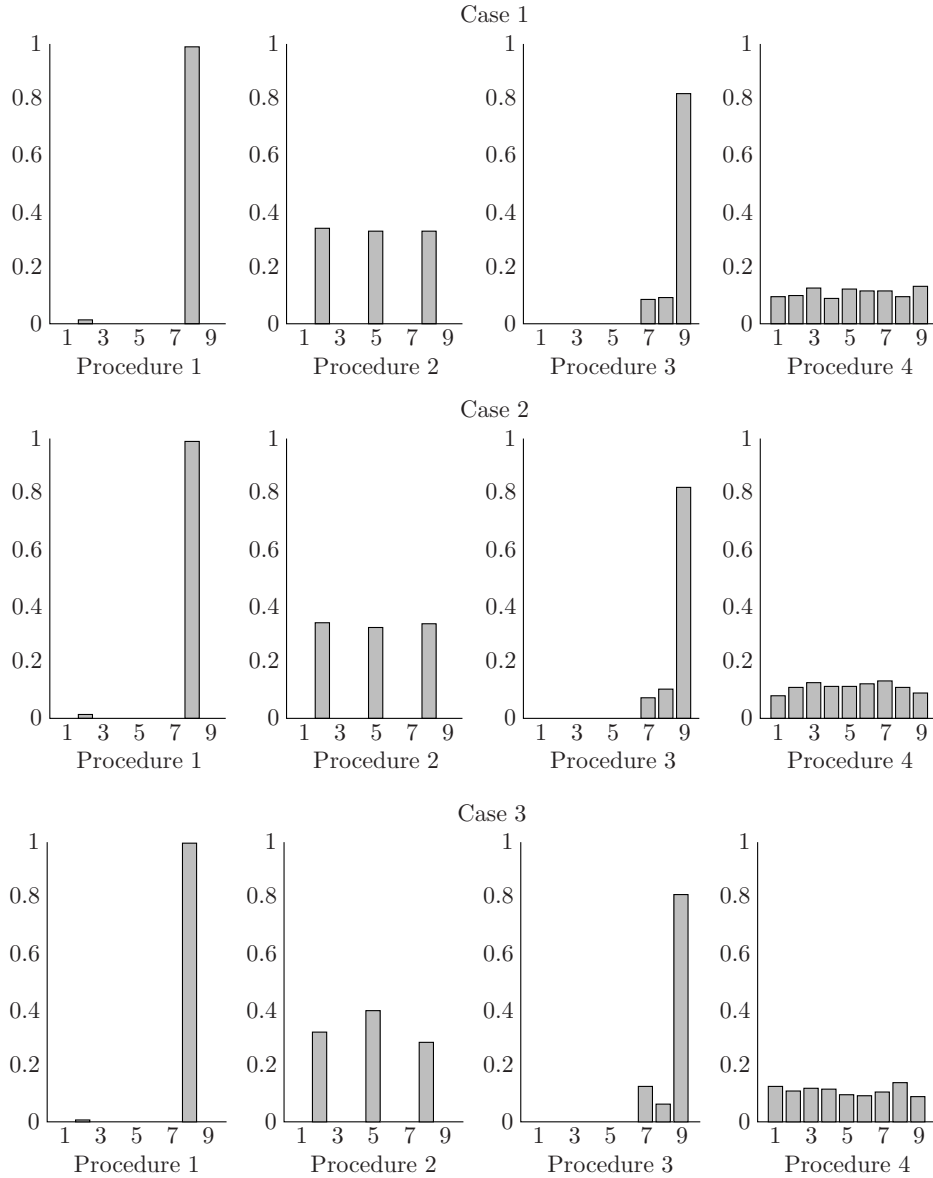
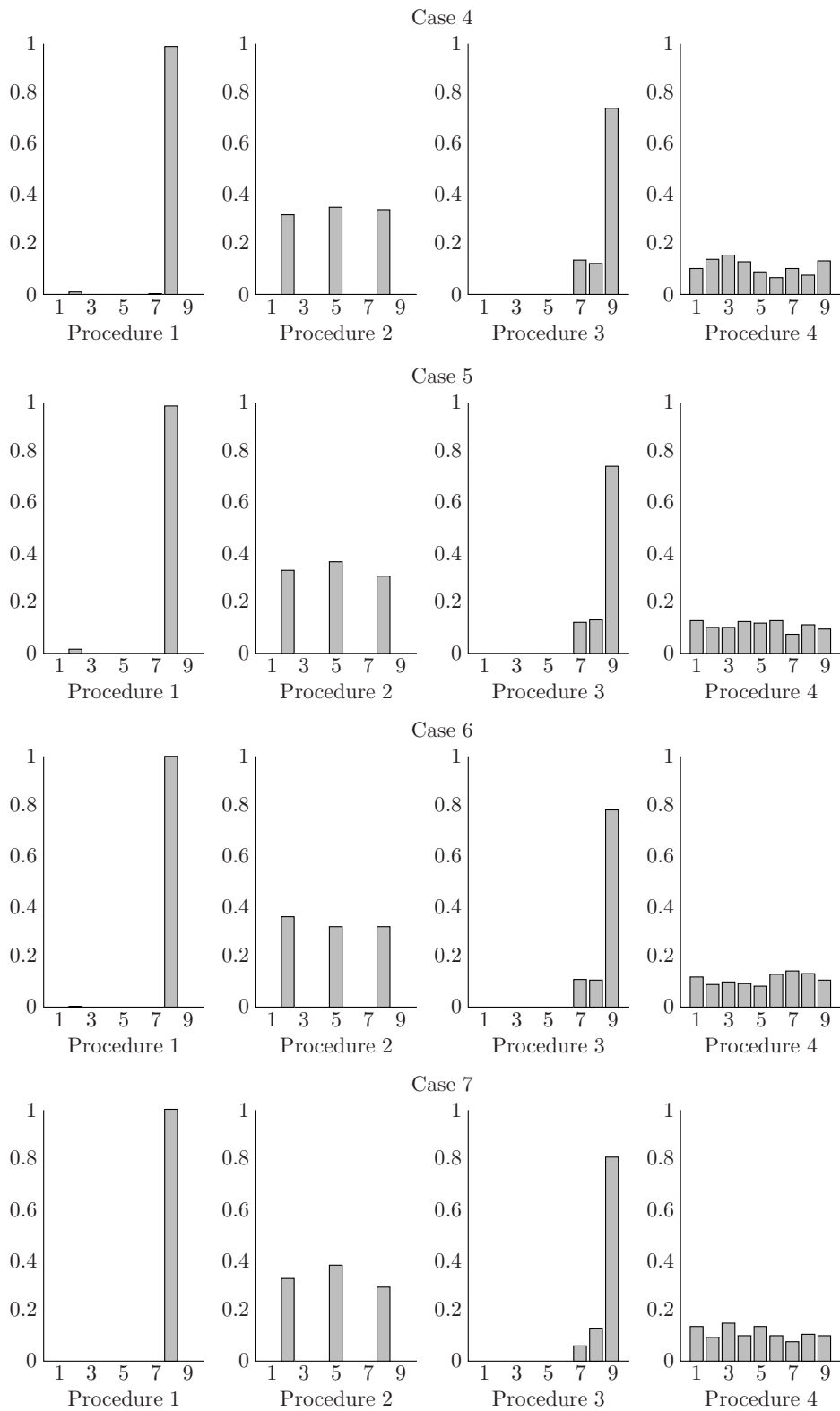**Figure 1:** *Choice of secondary series for Cases 1, 2 & 3.*

Figure 2:   *Choice of secondary series for Cases 4, 5, 6 & 7.*
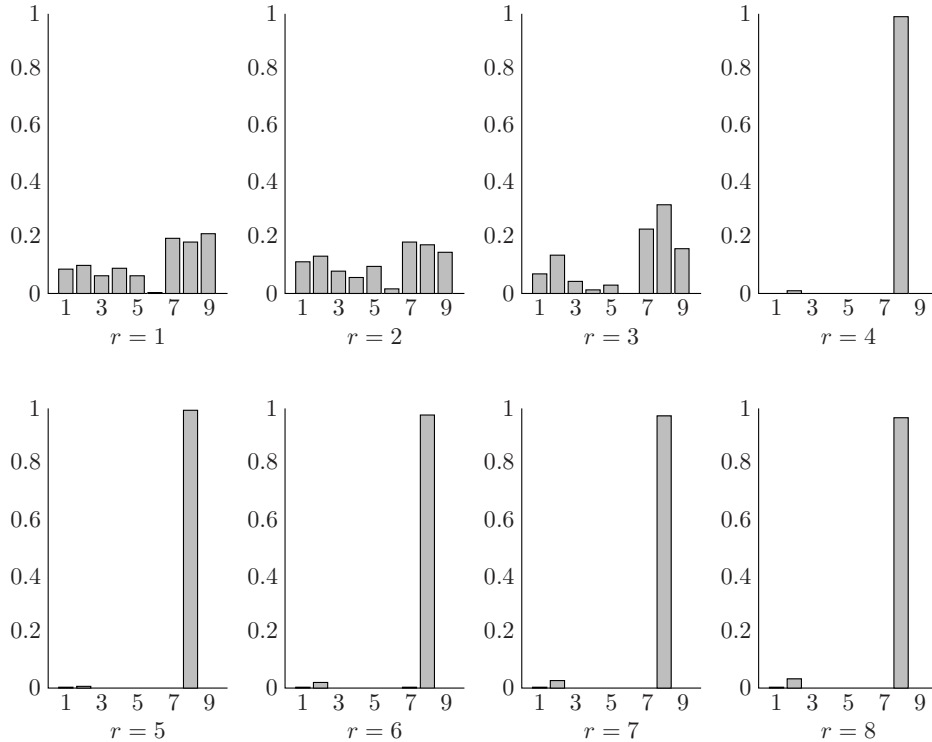
Case 4, Procedure 1

Figure 3:   *Choice of secondary series for different values of r.*

5   An example

In this section an application of the procedures to a practical dataset is described. The data are from Franses & Van Dijk (2013). The data consist of eight daily indices of stock markets in Amsterdam (EOE), Frankfurt (DAX), Hong Kong (Hang Seng), London (FTSE100), New York (S&P 500), Paris (CAC40), Singapore (Singapore All Shares) and Tokyo (Nikkei) for the period 6 January 1986, until 31 December 1997. The data from 1 January 1988 until 28 December 1997 are considered, giving a total of 2 606 observations. The study entailed 26 repetitions. In each repetition 100 consecutive observations were analysed using each of the five procedures and six step ahead forecasts were computed. The eight indices differ substantially in terms of their magnitudes, varying from approximately 200 up to approximately 22 000. Equation (4) was therefore modified to give a mean absolute relative percentage error measure,

$$\frac{1}{6} \sum_{t=N+1}^{N+6} \left| \frac{f_t - \hat{y}_t}{f_t} \right| \times 100.$$

In Tables 6 and 7 the results are summarized for $L = 24$ and $L = 48$, respectively.

The results in these tables were obtained by using the Amsterdam index as primary series with the other seven indices acting as candidate series. We see that all procedures

L = 24	Value of r								
	2	3	4	5	6	7	8	9	10
Proc 1	1.841	1.754	1.556	1.39	1.681	1.851	2.053	1.904	1.984
Proc 2	1.743	1.668	1.501	1.411	1.676	1.884	1.794	1.650	1.606
Proc 3	1.945	1.576	1.575	1.363	1.687	1.762	1.943	1.840	2.015
Proc 4	1.700	1.606	1.890	1.555	1.691	1.736	1.556	1.557	1.825
SSA	2.304	1.731	1.768	1.513	1.650	1.561	1.759	1.560	1.647

Table 6:   *Mean absolute relative percentage errors for $L = 24$.*

L = 24	Value of r								
	2	3	4	5	6	7	8	9	10
Proc 1	2.422	2.673	2.279	1.895	1.478	1.863	2.119	2.056	2.236
Proc 2	2.284	2.306	2.145	1.913	1.56	1.833	2.211	2.371	2.144
Proc 3	2.462	2.845	2.275	2.047	1.581	2.011	2.225	2.037	2.103
Proc 4	2.063	1.928	2.09	1.867	1.954	2.046	1.939	1.882	3.001
SSA	3.338	2.277	2.431	2.899	2.732	2.960	2.927	2.881	3.030

Table 7:   *Mean absolute relative percentage errors for $L = 48$.*

performed better for $L = 24$ than for $L = 48$. The best results were obtained for $r = 5$ with Procedure 3 giving the lowest error, closely followed by Procedure 1. We also found that the Singapore All Shares and Tokyo Nikkei indices were by far selected most frequently by Procedures 1 to 3. All in all it seems that some useful information could be provided by these indices when trying to predict the Amsterdam index.

6   Conclusions and questions for further research

In this paper the possibility of selecting a secondary time series from a set of available candidates for use in stepwise CSSA was investigated. Four procedures were studied reflecting different strategies to select the secondary series. These procedures were compared to using SSA where no secondary series is involved. It was found that selecting a secondary series with residuals negatively correlated to that of the primary series gave the best results, outperforming SSA in all cases. A question which deserves further investigation is how to select a value of r from the data.

Acknowledgement

References

[1] FLURY B, 1988, *Common principal components and related multivariate models*, Wiley, New York (NY).

[2] FRANSES PH & VAN DIJK D, 2000, *Non-linear time series models in empirical finance*, Cambridge University Press, Cambridge.

[3] FRANSES PH & VAN DIJK D, 2013, *Nonlinear Time Series Models in Empirical Finance*, [Online], [Cited: 24th October 2013], Available from `http://people.few.eur.nl/djvandijk/nltsmef/nltsmef.htm`.

[4] GOLYANDINA N, NEKRUTKIN VV & ZHIGLJAVSKY A, 2001, *Analysis of time series structure: SSA and related techniques*, Chapman & Hall, London.

[5] GOLYANDINA N & STEPANOV DV, 2005, *SSA-based approaches to analysis and forecast of multidimensional time series*, Proceedings of the 5th St. Petersburg Workshop in Simulation, Saint Petersburg, pp. 293–298.

[6] TRENDAFILOV NT, 2010, *Stepwise estimation of common principal components*, Computational Statistics and Data Analysis, **54(12)**, pp. 3446–3457.

[7] VENTER JH, 2000, *Time series analysis by unfolding*, Manuscript FABWI-N-WST: 2000-18, Centre for Business Mathematics and Informatics, Potchefstroom University, Potchefstroom.

[8] VILJOEN H & NEL DG, 2010, *Common singular spectrum analysis of several time series*, Journal of Statistical Planning and Inference, **140(1)**, pp. 260–267.

[9] VILJOEN H & STEEL SJ, 2012, *A comparison of common singular spectrum analysis and multi-channel singular spectrum analysis*, Technical report, University of Stellenbosch, Stellenbosch.