ICA time series analysis and independent component ordering

Lingxiao XU 119020061@link.cuhk.edu.cn

May 2021

1 ICA Model

Suppose that we are given an observed N time series $x_1(t)$, $x_2(t)$,...,$x_N(t)$, which are an instantaneous linear mixture of unknown mutually independent components $y_1(t)$, $y_2(t)$,...,$y_k(t)$. Thus, the observed series could be modeled in a matrix form:

$$X(t) = AY(t) \tag{1}$$

where $X(t) = [x_1(t), x_2(t), ..., x_N(t)]$ and $Y(t) = [y_1(t), y_2(t), ..., y_k(t)]$, and A is an $N \times K$ unknown transformation matrix. To extract independent component from the observed time signals, we need to find the de-mixing matrix W such that:

$$\hat{Y}(t) = WX(t) = WAY(t) \tag{2}$$

We used FastICA algorithm proposed by Aapo and Erkki (1) to estimate the de-mixing matrix W. The detailed algorithm is shown as follows:

1. Initialize the de-mixing matrix W
2. For each column vector w_p in W:

$$w^p = E[xg(w_p^T X)] - E[g^{'}(w_p^T x)]w \tag{3}$$

$$w_p = w_p - \sum_{j=1}^{p-1}(w_p^T w_j)w_j \tag{4}$$

$$w_p = \frac{w_p}{||w_p||} \tag{5}$$

3. If w_p not converge, repeat step 2.

After FastICA algorithm, the output is $W = [w_1, ..., w_K]$ and the resulting independent components is obtained by:

$$\hat{Y}(t) = W^T X(t) \tag{6}$$

2 Independent Component Ordering

The independent component selection method were introduced by Cheung and Xu in 2001 (2). Considering the kth time series $x_i(t)_{i=1}^N$, W_{ij}^{-1} could be considered as the contribution of component $\hat{y_j}$ to the reconstruction of x_i:

$$\hat{u}_{ij}(t) = W_{ij}^{-1}\hat{y_j}(t) \tag{7}$$

The resconstruction of x_i by using the first m independent components under a give list L_i is therefore given by:

$$x_{L_i}^{\hat{m}}(t) = \sum_{r=1}^{m} u_{i\hat{l(r)}}(t) \tag{8}$$

where $l(r)$ is the rth element of L_i. To compare the reconstructed signal with the original one, we denote $Q(x_i, x_{L_i}^{\hat{m}})$ to be the corresponding reconstruction error. Thus, the cumulative data construction error is given by:

$$J_{L_i} = \sum_{m=1}^{k} J_{L_i}(m) = \sum_{m=1}^{k} Q(x_i, x_{L_i}^{\hat{m}}) \tag{9}$$

Cheung and Xu used the Relative Hamming Distance to measure the the reconstruction error based on the consideration of the trend of a time series may be mostly controlled by the underlying independent components:

$$Q(x_i, \hat{x_i}^m) = RHD(x_i, \hat{X_{L_i}}) = \frac{1}{N-1}\sum_{t=1}^{N-1}[R_i(t) - R_{L_i}^{\hat{m}}(t)] \tag{10}$$

where

$$R_i(t) = sign[x_i(t+1) - x_i(t)], R_{L_i}^{\hat{m}}(t) = sign[x_{L_i}^{\hat{m}}(t+1) - x_{L_i}^{\hat{m}}(t)] \tag{11}$$

and

$$sign(r) = \begin{cases} 1 & ifr > 0 \\ 0 & ifr = 0 \\ -1 & ifr < 0 \end{cases} \tag{12}$$

Hence, the optimum order list L_i^* is given by:

$$L_i^* = argmin_{L_i} J_{L_i} \tag{13}$$

Since the exhaustive permutation of L_i^* is time-consuming. We considered Testing-and-Acceptance algorithmTnA) to select the independent components. The basic procedure of TnA is: from the set of k independent components, find one independent component that gives the minimum reconstruction error after removing it. Remove the resulting independent component. Repeat the above process until the desired result.

3 Performing ICA and ordering on the stock index data set

This paper uses the SSE Composite Index (000001.SS) and Shenzhne Component Index (399001.SZ) from 2019 to 2021, which covers two stock market in China. The data set contains 5 categories: close price, the highest price, the lowest price, and trade volume. Given that ICA has a better performance on the stationary time series, we pre-processed the data set by:

$$x(\hat{t}) = log(x(t+1)) - log(x(t)) \tag{14}$$

$$\widetilde{x(t)} = \frac{x(\hat{t}) - min(\hat{(}x)}{max(\hat{x}) - min(\hat{x})} \tag{15}$$

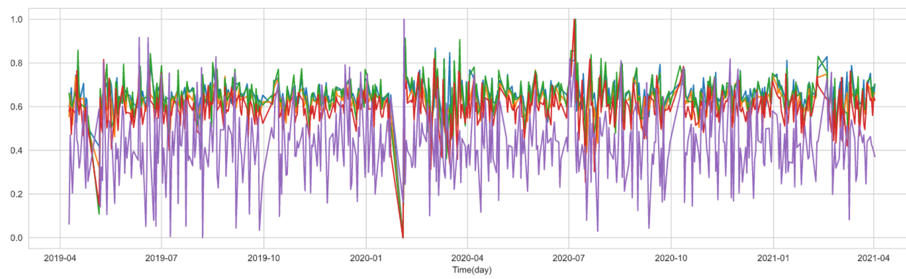


Figure 1: Mixture Signals of 000001.SS

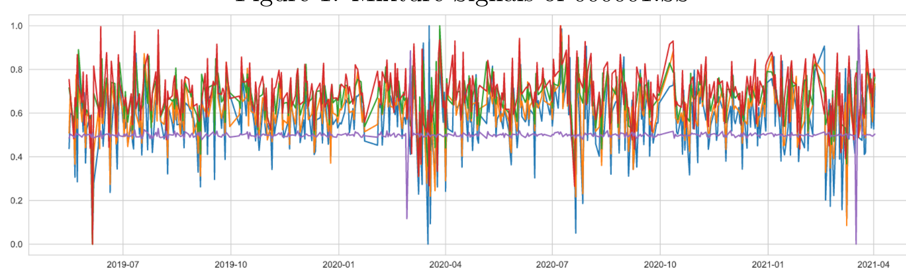
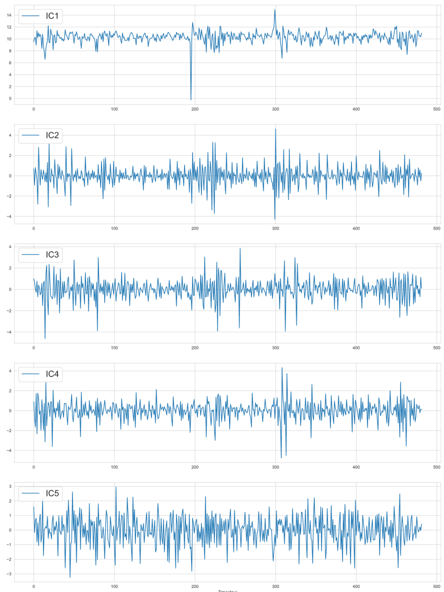


Figure 2: Mixture Signals of 399001.SZ

Then, we perform ICA on two stock indexes respectively:



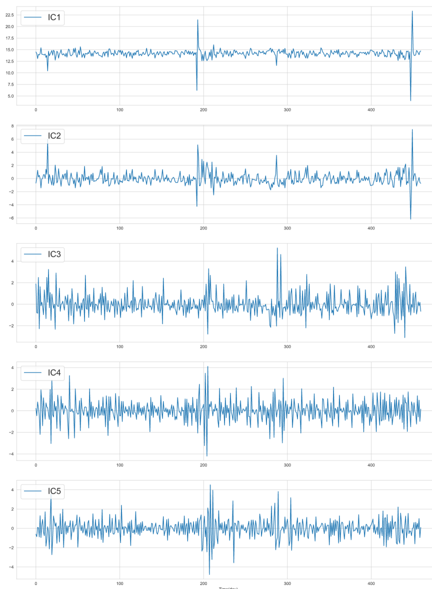


Figure 3: Independent components of 000001.SS

Figure 4: Independent components of 399001.SZ

Finally, we perform independent component ordering on the resulting ICs.Since the small number of obtained ICs, we only performs TnA algorithm for one iteration.

Table 1: Ordering of independent components

	000001.SS	399001.SZ
IC1	8.257	9.699
IC2	7.076	8.671
IC3	7.941	9.255
IC4	8.091	8.427
IC5	5.987	9.159

4 References

References

[1] A. Hyvärinen & E. Oja, Independent component analysis: algorithms and applications, Neural Networks, Volume 13, Issues 4–5, 2000, Pages 411-430, ISSN 0893-6080, https://doi.org/10.1016/S0893-6080(00)00026-5.

[2] Yiu-ming Cheung, Lei Xu, Independent component ordering in ICA time series analysis, Neurocomputing, Volume 41, Issues 1–4, 2001, Pages 145-152, ISSN 0925-2312, https://doi.org/10.1016/S0925-2312(00)00358-1.