Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Composite Index Construction with Expert Opinion

Rong Chen, Yuanyuan Ji, Guolin Jiang, Han Xiao, Ruoqing Xie & Pingfang Zhu

To cite this article: Rong Chen, Yuanyuan Ji, Guolin Jiang, Han Xiao, Ruoqing Xie & Pingfang Zhu (2023) Composite Index Construction with Expert Opinion, Journal of Business & Economic Statistics, 41:1, 67-79, DOI: 10.1080/07350015.2021.2000418

To link to this article: https://doi.org/10.1080/07350015.2021.2000418

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 1, 67–79
https://doi.org/10.1080/07350015.2021.2000418

Composite Index Construction with Expert Opinion

Rong Chen[a], Yuanyuan Ji[b], Guolin Jiang[b], Han Xiao[a], Ruoqing Xie[b,c], and Pingfang Zhu[b,c]

[a]Department of Statistics, Rutgers University, Rutgers, NJ; [b]Research Center of Econometrics, Shanghai Academy of Social Sciences, Shanghai, PR China; [c]Institute of Economics, Shanghai Academy of Social Sciences, Shanghai, PR China

ABSTRACT

Composite index is a powerful and popularly used tool in providing an overall measure of a subject by summarizing a group of measurements (component indices) of different aspects of the subject. It is widely used in economics, finance, policy evaluation, performance ranking, and many other fields. Effective construction of a composite index has been studied extensively. The most widely used approach is to use a linear combination of the component indices, where the combination weights are determined by optimizing an objective function. To maximize the overall variation of the resulting composite index, the combination weights can be obtained through principal component analysis. In this article, we propose to incorporate expert opinions into the construction of the composite index. It is noted that expert opinion often provides useful information in assessing which of the component indices are more important for the overall measure of the subject. We consider the case that a group of experts have been consulted, each providing a set of importance scores for the component indices, along with a set of confidence scores which reflects the expert's own confidence in his/her assessment. In addition, the constructor of the composite index can also provide an assessment of the expertise level of each expert. We use linear combinations to construct the composite index, where the combination weights are determined by maximizing the sum of resulting composite index variation and the negative weighted sum of squares of deviation between the combination weights used and the experts' scores. A data-driven approach is used to find the optimal balance between the two sources of information. Theoretical properties of the procedure are investigated. Simulation examples and an economic application on constructing science and technology development index is carried out to illustrate the proposed method.

ARTICLE HISTORY
Received July 2020
Accepted October 2021

KEYWORDS
Composite index; Expert opinion; Factor model; Principal component analysis

1. Introduction

Composite index is used to provide a summary measurement of a complex subject with many different features. By measuring the features separately as the component indices, then numerically combining them into a single value as the composite index, it is often used for comparison and ranking. It is widely used in economics, finance, policy evaluation, performance ranking, and many other fields. For example, the Leading Economic Index (LEI) published by The Conference Board is composed of 10 economic component indices whose change tend to precede changes in the overall economy (Stock and Watson 1989). Market indices such as S&P 500 Index and Dow Jones Index are composite indices by combining the prices of a group of stocks (Cross 1973; Kawaller, Koch, and Koch 1987). Volatility index is a composite index constructed using information in option prices of different strikes and expiration dates (Whaley 2008). The *Regulatory Indicators Assessment* index, the *stakeholder engagement* index and the *ex post evaluation* index constructed by OECD are all composite indices to evaluate the regulatory policy and governance (Kaur and Lodhia 2014). College rankings are done by using a composite index combining various aspects of the university, including student life, graduation rate, funding levels to faculty research ability (Karabel and Astin 1975).

The research on how to construct a composite index is vast. A common approach uses a linear combination of the individual observed component indices to construct the composite index. With such an approach, the main task is then to determine the combination weights. Two methods, the objective weighting method and the subjective weighting method, are typically used, based on the information being used. The objective weighting method relies only on the measured data and is widely studied, including the principal component analysis (PCA) approach (Alzate and Suykens 2010; Tavoli et al. 2013), the entropy weighting method (Hoskisson et al. 1993; Jing, Ng, and Huang 2007; Chen and Li 2010; Shemshadi et al. 2011), the clustering approach (Milligan 1989; Eisen et al. 1998; Yu, Yang, and Lee 2011) and others. The subjective weighting methods include ranking weighting method (Roszkowska 2013), analytic hierarchy process (AHP) (Al-Harbi 2001) and others. These subjective weighting method heavily depends on the experts' professionalism. On the other hand, the objective weighting method, which determines the combination weights solely based on the observed component indices, along with certain mathematical models and assumptions, neglects the subjective judgment information of the decision makers and may result in misleading and counter intuitive results (Aalianvari, Katibeh, and Sharifzadeh 2012). There are some recent

CONTACT Rong Chen rongchen@stat.rutgers.edu Department of Statistics, Rutgers The State University of New Jersey, New Brunswick, NJ 08901-8554.

researches on multi-criteria decision making (MCDM) methods (Zardari et al. 2015), such as the Best-Worst Method (Rezaei 2015), Analytic Network Process (Saaty 2008; Meade and Presley 2002), and context tree weighting (Willems, Shtarkov, and Tjalkens 1995; Garivier 2006). Those MCDM methods are data-driven, and are used in many fields including systems engineering studies (Kujawski 2003).

In this article, we develop a novel method that combines the objective information (data) with subjective information (expert opinion). By effectively combining both sources of information when available, it makes the construction more accurate and reduces biases introduced by either source of information. Specifically, we adopt a factor model setting to use the observed component indices, and use a least-square penalty term to incorporate the expert's opinion regarding the combination weight (importance) of each component index, along with a self-assessment of confidence from the experts and an expertise score from the composite index constructor. Such a comprehensive collection of subjective information allows for diverse opinions and different levels of expertise on different subjects. We use a penalty parameter to balance the influence of the objective information and subjective information. It can be viewed as a ratio of the noise levels in the data and in expert opinions. The optimal penalty parameter can be obtained by maximizing the overall accuracy, through a cross-validation approach.

The rest of the article is organized as follows. In Section 2, we introduce the factor model assumed for the observed component indices, and the data structure of the experts' opinion along with their confidence scores. We then introduce the objective function that combines both sets of information. The composite index is constructed by finding a set of combination weights that optimizes the objective function. In Section 3, we investigate the theoretical properties of the construction. Finite sample properties of the developed procedure are investigated in Section 4 through a simulation study. An economic application on constructing a composite index on science and technology development is shown in Section 5. Section 6 concludes.

2. Data, Model Setting, and Construction Procedure

2.1. Data and Model Setting

Suppose we have K candidate component indices $\{x_{ki}\}$ for $k = 1, \ldots, K$ with N observations $i = 1, \ldots, N$, to be included in the construction of the composite index. We will use a linear combination of the component indices to construct the composite index. Specifically, the composite index is in the form

$$C_i = \sum_{k=1}^{K} w_k x_{ki}, \tag{1}$$

where the combination weight $\boldsymbol{w} = (w_1, \ldots, w_K)'$, normalized so that $||\boldsymbol{w}||_2^2 = \sum_{k=1}^{K} w_k^2 = 1$, needs to be determined.

Traditional composite index construction using PCA approach (Li et al. 2012; Nardo et al. 2008) finds the combination weight $\boldsymbol{w}$ so that the resulting component indices have the largest variance among all possible such linear combinations – the first principle component. Specifically, let $\hat{\boldsymbol{w}}$ be the normalized eigenvector corresponding to the largest eigenvalue of $\hat{\Sigma}_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i'$, the sample covariance matrix of $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{Ki})'$. That is,

$$\hat{\boldsymbol{w}} = \arg \max_{||\boldsymbol{w}||=1} \boldsymbol{w}' \hat{\Sigma}_N \boldsymbol{w}. \tag{2}$$

The composite index is then constructed as $\hat{f}_i = \hat{\boldsymbol{w}}' \boldsymbol{x}_i$. The variance of $\hat{f}_i$ is the largest among all possible such combinations by the construction of $\hat{\boldsymbol{w}}$ in Equation (2). PCA estimation is usually done using singular value decomposition, though the original optimization formulation can be useful when additional information is used as it allows modifications of the objective function.

We note that the solution of the PCA approach above is the same as fitting a single-factor model

$$\boldsymbol{x}_i = \boldsymbol{w} f_i + \boldsymbol{\epsilon}_i,\ k = 1, \ldots, K, i = 1, \ldots, N, \tag{3}$$

where $\boldsymbol{w} = (w_1, \ldots, w_K)$ is the loading vector, and f_i is the latent common factor. The noise $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \ldots, \epsilon_{Ki})'$ is assumed to be iid with zero mean and covariance matrix Σ_ϵ. The estimator of the latent factor f_i is $\hat{\boldsymbol{w}}' \boldsymbol{x}_i$, under a general condition on $\mathrm{Var}(f_i)$ and Σ_ϵ.

In addition to observing the K component indices, we also assume that we have surveyed total J experts who have provided their direct assessments of the combination weight $\boldsymbol{w}$ for the construction of the composite index, along with a confidence score on each of the combination weights provided. Let (s_{kj}, γ_{kj}), $k = 1, \ldots, K$, $j = 1, \ldots, J$, be the importance score and its corresponding confidence score provided by the jth expert. The importance score to each component index reflects the experts' opinion on how much combination weight should be assigned to a candidate component index in the construction. The score s_{kj} is normalized so that $\sum_{k=1}^{K} s_{kj}^2 = 1$. We will assume that the experts' scores are *proportionally unbiased*, with $\mathbb{E}[s_{kj}] = \delta w_k$ where w_k are given in Equation (3) and $\delta > 0$ is a scalar. The reason for the proportional unbiasedness assumption used here instead of simple unbiasedness assumption is due to the fact that the two conditions $\sum_{k=1}^{K} s_{kj}^2 = 1$ and $\sum_{k=1}^{K} w_k^2 = 1$ make the simple unbiasedness assumption impossible.

The confidence score γ_{kj} reflects the expert's assessment on his/her own expertise level on the subject, possibly with different levels of expertise among the K component indices. The larger the γ_{kj} is, the higher the experts' confidence is on the k's component. If the expert knows one component index very well, then he/she will assign a large confidence score. Otherwise, a small score will be assigned. Jiang, Liu, and Zhu (1996) considered the situation that an expert will provide a range (interval) for each of the combination weights. Corresponding to our setting, the center point of their interval would be the importance score and the inverse of the interval width would be the confidence score in our case. In this article, the confidence score γ_{kj} is restricted to $[0, 1]$.

Furthermore, the constructor of the composite index may assign an "expertise" score c_j to the jth expert. This provides a ranking among the experts in terms of their relative expertise on the construction of the index of interest. We restrict the value of c_j to $(0, 1]$.

It is the aim of this article to construct the composite index by combining both the observed component indices and the expert

opinion. Statistically speaking, our construction is based on a model with two parts: a factor model on the observed indices, and IID scores which are "proportionally unbiased" from the experts. To use both sources of information, we use a combined least-square objective function in the form

$$\begin{aligned} g^*_{N,J}(\boldsymbol{w},\delta) &= \boldsymbol{w}'\textstyle\sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{w} - Q\sum_{j=1}^J c_j \\ &\quad \left\{\textstyle\sum_{k=1}^K \gamma_{kj}(s_{kj}-\delta w_k)^2\right\} \\ &= \boldsymbol{w}'\textstyle\sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{w} - Q\delta^2\boldsymbol{w}'\sum_{j=1}^J \Gamma_j\boldsymbol{w} \\ &\quad +2Q\delta\textstyle\sum_{j=1}^J \boldsymbol{s}_j'\Gamma_j\boldsymbol{w} - Q\sum_{j=1}^J \boldsymbol{s}_j'\Gamma_j\boldsymbol{s}_j. \end{aligned} \tag{4}$$

subject to $0 \le w_k \le 1$ and $||\boldsymbol{w}||_2^2 = 1$, and $\delta > 0$ is a simple scalar, where $\Gamma_j = \text{diag}(c_j\gamma_{1j},\ldots,c_j\gamma_{Kj})$.

The first term in Equation (4) is the original criterion of estimating the optimal combination weight using PCA without expert input. It is a classical quadratic maximization problem. The second term is the weighted least-square term for fitting the expert opinions, adjusted by their confidence scores and expertise scores.

The constant Q is a penalty parameter which balances the variance of the linear combination in the first term and the error variance in fitting the expert opinions in the second term. It is an important parameter. When Q is large, the objective function $g^*_{N,J}(\boldsymbol{w})$ in Equation (4) puts more combination weights on the second term related to the expert opinion. Hence, the solution $\hat{\boldsymbol{w}}_{N,J}$ would be closer to the optimizer of the second term. Similarly, when Q is small, the solution would be closer to the PCA solution that maximizes only the first term, without the expert opinions. In fact, the optimal Q should reflect the comparison between the noise level in the observed data and the noise level in the expert opinion. When the noise level in the observed data is larger than that in the expert opinion, we would trust the experts more, hence using larger Q.

Remark 1. If we assume normality on $\boldsymbol{x}_i$ and $\boldsymbol{s}_j$, then it is also possible to estimate $\boldsymbol{w}$ using the maximum likelihood method. In this article, we choose to use the weighted least-square criterion so that we do not need to specify and estimate the error covariance matrix.

Remark 2. If we treat the expert information as prior information, then a Bayesian approach can be used as well. However, it would require to specify the expert score distribution as well as noise distribution in the factor model. We do not use the Bayesian approach in this article.

Optimizing $g^*_{N,J}(\boldsymbol{w})$ in Equation (4) is equivalent to optimize

$$g_{N,J}(\boldsymbol{w},\delta) = a_{N,J}\boldsymbol{w}'\hat{\Sigma}_N\boldsymbol{w} - b_{N,J}Q\delta^2\boldsymbol{w}'\bar{\Gamma}_J\boldsymbol{w} + 2b_{N,J}Q\delta\bar{\boldsymbol{s}}_J'\boldsymbol{w}, \tag{5}$$

where $\hat{\Sigma}_N = \frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{x}_i'$ is the sample covariance matrix of $\boldsymbol{x}_i$, $\bar{\boldsymbol{s}}_J = \frac{1}{J}\sum_{j=1}^J \Gamma_j\boldsymbol{s}_j$ is the weighted average of the expert scores $\boldsymbol{s}_j$, and $\bar{\Gamma}_J = \frac{1}{J}\sum_{j=1}^J \Gamma_j$. The constants $a_{N,J} = N/(N+J)$ and $b_{N,J} = J/(N+J)$ reflect sample proportions from the two sources of information.

The estimator of $\boldsymbol{w}$ is then

$$\hat{\boldsymbol{w}}_{N,J} = \arg\max_{||\boldsymbol{w}||_2^2=1}\max_{\delta} g_{N,J}(\boldsymbol{w},\delta). \tag{6}$$

Remark 3. Note that if we do not have the observed data, then $\boldsymbol{s}_* = \bar{\Gamma}_J^{-1}\bar{\boldsymbol{s}}_J$ would be the solution of the second term in Equation (4) without the $\boldsymbol{w}'\boldsymbol{w} = 1$ constraint. It provides a summary of the expert scores, adjusted by the confidence scores and expertise scores. In particular, if all γ_{kj} are the same, then $\bar{\Gamma}_J$ is in a form of a scalar matrix. Then the solution to the second term in Equation (4) with the $\boldsymbol{w}'\boldsymbol{w} = 1$ constraint would be $\bar{\boldsymbol{s}}_J/\sqrt{\bar{\boldsymbol{s}}_J'\bar{\boldsymbol{s}}_J}$, a normalized average of the expert scores.

Solution to Equation (5) can be obtained through quadratic programming, under quadratic equality constraints. We also note that the objective function is a difference of two convex functions in a constrained space. Optimization of such a function is easy and fast, with good properties (Markowitz 1956).

Once we obtain $\hat{\boldsymbol{w}}_{N,J}$ through optimizing (5), the composite index can be constructed with $C_i = \hat{\boldsymbol{w}}_{N,J}'\boldsymbol{x}_i$ and the fitted value of $\boldsymbol{x}_i$ can be obtained with $\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{w}}_{N,J}\hat{\boldsymbol{w}}_{N,J}'\boldsymbol{x}_i$.

2.2. Geometry Interpretation

The quadratic term involving $\boldsymbol{w}$ in Equation (5) can be written as $\boldsymbol{w}'\Sigma_*\boldsymbol{w}$, where $\Sigma_* = a_{N,J}\hat{\Sigma}_N - b_{N,J}Q\delta^2\bar{\Gamma}_J$. Therefore, different from the PCA approach in which the covariance matrix is always positive semidefinite, the combined objective function $g_{N,J}(\boldsymbol{w},\delta)$ is in a quadratic form with the "covariance" matrix Σ_*, which can be positive definite, negative definite or indefinite, depending on the penalty parameter Q. When Q is small, the matrix Σ_* is more likely to be positive definite; when Q is very large, the matrix would be negative definite.

To illustrate, suppose $K = 2$. The surface in Figure 1 shows the quadratic function $g_{N,J}(\boldsymbol{w},\delta)$ for a fixed δ. The unit circle constraint of $\boldsymbol{w}$ is represented by the cylinder space. The constraints restrict the quadratic maximization problem in a lower dimensional constrained space which is also compact. Note that there is no "corners" in the lower dimensional space, hence the function $g_{N,J}(\boldsymbol{w},\delta)$ is also continuous in the reduced space, hence an optimization solution always exists, no matter whether the matrix Σ_* is positive definite, negative definite, or indefinite.

2.3. Determination of the Penalty Parameter

The penalty parameter Q is an important tuning parameter in practice. It has significant impact on the constructed composite index as discussed in Section 2.1. Here, we propose to use a combined M-fold ($M = M_1M_2$) cross-validation for its determination in practice. Specifically, the original observed sample is randomly partitioned into M_1 equal size subsets $D_1,\ldots,D_{M_1}$. The experts scores are divided into M_2 equal size subsets $D^*_1,\ldots,D^*_{M_2}$. We use $M_1 - 1$ observed data sample subsets and $M_2 - 1$ experts scores subsets as the training data for estimating $\boldsymbol{w}$. Then $\hat{\boldsymbol{w}}_{N,J}$ is used to predict the validation subset under the factor model setting and the prediction sum of squares of errors is obtained. Specifically, define

$$\begin{aligned} \text{CV}(Q) = \frac{1}{M_1M_2}\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\Bigg[&-\sum_{i\in D_{m_1}}[\boldsymbol{x}_i'\hat{\boldsymbol{w}}^{(-m)}_{N,J}(Q)]^2 \\ &+ C\sum_{j\in D^*_{m_2}}||\boldsymbol{s}_j - \hat{\delta}^{(-m)}(Q)\hat{\boldsymbol{w}}^{(-m)}_{N,J}(Q)||_2^2\Bigg], \end{aligned} \tag{7}$$

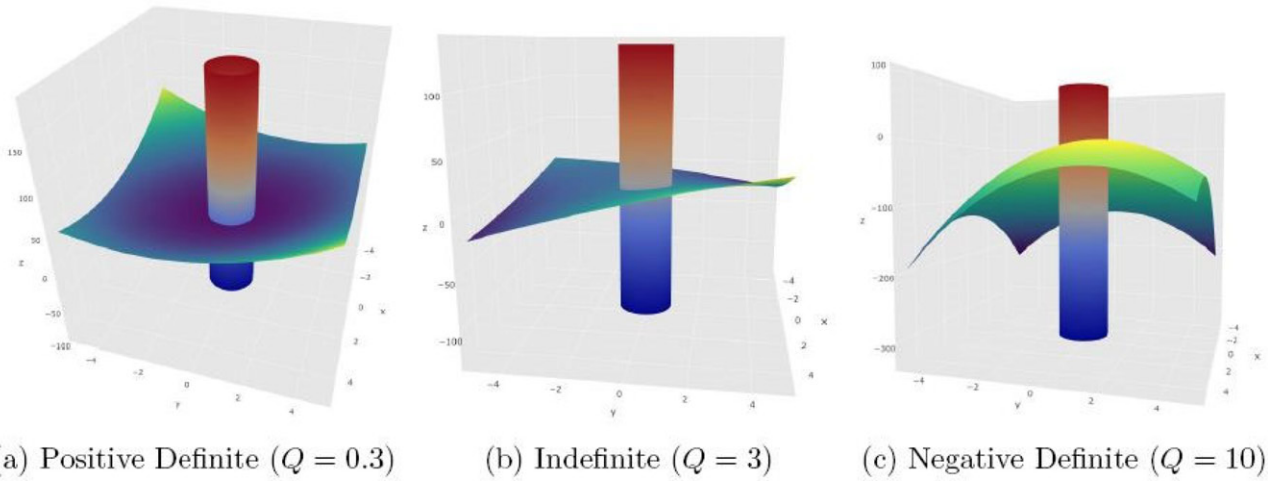

Figure 1. Estimation picture.

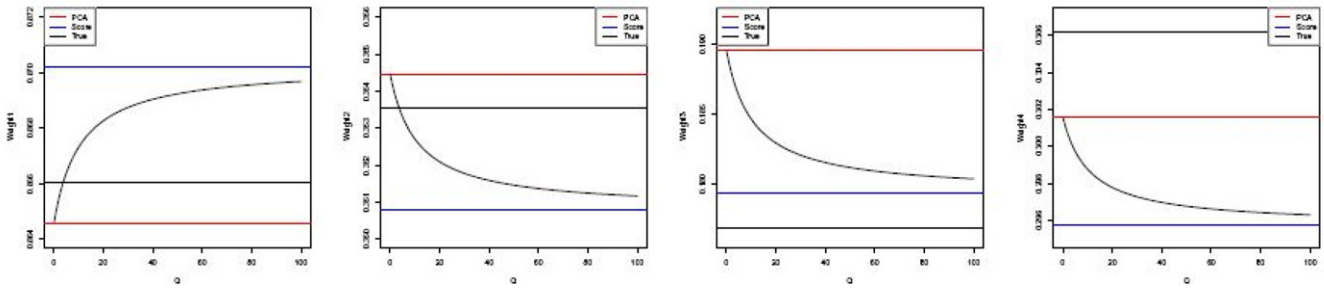

Figure 2. Solution path of $\hat{\boldsymbol{w}}$ as a function of Q.

Table 1. Estimated weights in Exercise (i).

Estimation	w_1	w_2	w_3	w_4
True Weight	0.866025	0.353553	0.176777	0.306186
PCA($Q = 0$)	0.864550	0.354462	0.189595	0.301601
$Q = 100$	0.869681	0.351170	0.180353	0.296321
Score($Q = \infty$)	0.870224	0.350812	0.179361	0.295753

where $\hat{\boldsymbol{w}}_{N,J}^{(-m)}(Q)$ is the optimal $\boldsymbol{w}$ estimated using Q and without using data in D_{m_1} and $D_{m_2}^*$. Optimal Q is the one with the smallest cross-validation error $\mathrm{CV}(Q)$. Here C is a tuning parameter that balanced the two sources of errors. For small sample cases, leave-one-out cross-validation is used.

3. Large-Sample Properties

In this section, we investigate the large-sample properties of the estimator proposed in the preceding sections. We consider the rate of convergence when the number of involved component indices K is fixed or grows with the sample sizes N and J. In addition, we also establish the central limit theorem of $\hat{\boldsymbol{w}}$ when K is fixed. In the high dimensional setting, we let K and J be constants depending implicitly on N, and consider the asymptotics as $N \to \infty$.

We first list the assumptions for the fixed dimensional case. For the rest of this article, we use $||\cdot||$ to denote the spectral norm of a matrix, and the Euclidean norm of a vector. The symbol $\Rightarrow$ denotes the convergence in distribution.

Assumption 1. Assume $K > 0$ is fixed, and $\boldsymbol{x}_i$ are iid with mean zero and covariance matrix Σ_0. Let λ_1 be the largest eigenvalue of Σ_0, and $\boldsymbol{w}_0$ the corresponding eigenvector. Assume that the second largest eigenvalue λ_2 of Σ_0 is strictly smaller than λ_1. We also assume that $\mathrm{var}(\boldsymbol{w}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{w})$ is bounded for all $\boldsymbol{w}$ with $||\boldsymbol{w}||_2^2 = 1$.

Assumption 2. The expert scores $\boldsymbol{s}_j$ are independent, satisfying $||\boldsymbol{s}_j|| = 1$, $\mathbb{E}(\boldsymbol{s}_j) = \delta_0\boldsymbol{w}_0$, and $\mathrm{var}(\boldsymbol{s}_j) = \Sigma_s$.

Assumption 3. The observed data $\{\boldsymbol{x}_i,\ i = 1, 2, \ldots, N\}$ and the expert scores $\{\boldsymbol{s}_j,\ j = 1, 2, \ldots, J\}$ are independent.

Assumption 4. Let $\bar{\Gamma}_J = \frac{1}{J}\sum_{j=1}^J \Gamma_j$. Assume that $\bar{\Gamma}_J \to \Gamma_0$, and $J^{-1}\sum_{j=1}^J \Gamma_j\Sigma_s\Gamma_j' \to \tilde{\Sigma}_s$ as $J \to \infty$, where Γ_0 is a constant diagonal matrix, with positive diagonal elements, and $\tilde{\Sigma}_s$ is a constant positive-definite matrix.

Assumption 5. Assume $N/(N + J) \to a$ as $N, J \to \infty$, where $0 \le a \le 1$.

For the high-dimensional case, we replace Assumptions 1 and 4 with the following:

Assumption 1(*): We assume $K \to \infty$ as $N, J \to \infty$. Assume that $\boldsymbol{x}_i$ are iid with mean zero and covariance matrix Σ_{0N}. Let λ_{1N} be the largest eigenvalue of Σ_{0N}, and $\boldsymbol{w}_{0N}$ the corresponding eigenvector. Let λ_{2N} be the second largest eigenvalue of Σ_{0N}, assume $\liminf_{N\to\infty}(\lambda_{1N} - \lambda_{2N}) > 0$. We consider the optimization problem (5) with a covariance matrix estimator $\check{\Sigma}_N$

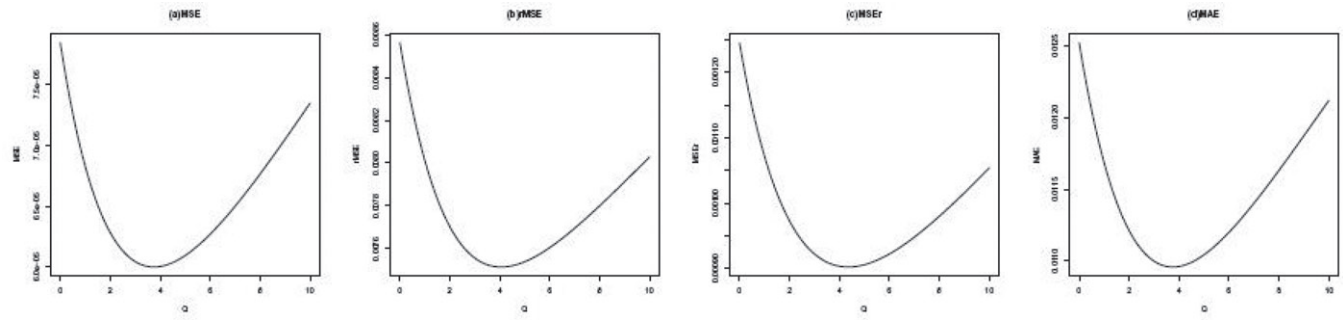

Figure 3. Performance measure as function of Q for $(N, J) = (400, 40)$ case.

Table 2. Optimal Q and its corresponding performance.

	MSE $\times 10^{-4}$	Q	RMSE $\times 10^{-2}$	Q	MSEr $\times 10^{-3}$	Q	MAE	Q
$N = 100, J = 10$	2.50	3.4	1.53	3.5	4.34	4.3	0.022	3.4
$N = 100, J = 40$	1.32	3.8	1.08	5.3	1.90	4.2	0.016	3.8
$N = 400, J = 10$	0.78	2.9	0.86	2.8	1.22	3.5	0.012	2.9
$N = 400, J = 40$	0.60	3.7	0.75	4.1	0.90	4.4	0.011	3.7

Table 3. Optimal Q under various noise levels.

		σ_s									
	Q	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	0.1	4.0	1.0	0.5	0.5	0.3	0.3	0.01	0.0	0.01	0.3
	0.2	20.0	3.0	2.0	1.0	0.7	0.01	0.4	0.4	0.5	0.3
	0.3	40.0	8.0	3.0	2.0	1.0	2.0	0.4	0.7	0.5	1.0
	0.4	100.0	10.0	8.0	5.0	4.0	2.0	2.0	1.0	2.0	1.0
σ_ϵ	0.5	100.0	20.0	10.0	9.0	5.0	5.0	2.0	2.0	2.0	1.0
	0.6	100.0	30.0	20.0	10.0	8.0	5.0	5.0	3.0	3.0	3.0
	0.7	100.0	80.0	20.0	20.0	20.0	10.0	5.0	5.0	4.0	4.0
	0.8	100.0	100.0	40.0	20.0	10.0	10.0	9.0	5.0	6.0	5.0
	0.9	100.0	100.0	50.0	50.0	20.0	20.0	10.0	10.0	8.0	9.0
	1.0	100.0	100.0	90.0	40.0	30.0	30.0	20.0	10.0	10.0	10.0

satisfying $||\check{\Sigma}_N - \Sigma_{0N}|| = O_P(\Delta_{1N})$, where $\Delta_{1N} \to 0$ as $N \to \infty$.

Assumption 4(*): Assume that the smallest diagonal element of $\bar{\Gamma}_J$ is positive and bounded away from zero.

The following remarks provide some comments on the assumptions.

Remark 4. Assumption 1 is typical for principle component analysis. The iid assumption of $\{\boldsymbol{x}_i\}_{i=1}^N$ can be relaxed. The results still hold if $\{\boldsymbol{x}_i\}_{i=1}^N$ satisfy certain mixing conditions. A more widely used but more restricted assumption (typical for a factor model) that $\Sigma_0 = (\lambda_1 - \lambda_2)\boldsymbol{w}_0\boldsymbol{w}_0' + \lambda_2 \boldsymbol{I}$ can be used here as well, as it also guarantees that $\boldsymbol{w}_0$ maximizes $\boldsymbol{w}'\Sigma_0\boldsymbol{w}$. Here we choose to allow the noise term $\boldsymbol{x}_i - \boldsymbol{w}_0\boldsymbol{w}_0'\boldsymbol{x}_i$ to have nonzero correlation.

Remark 5. Assumption 2 assumes that all experts make their assessments independently. Assumption 3 assumes that the experts do not make their assessments based on the observed data.

Remark 6. Assumption 4 is needed to derive the central limit theorem for fixed K. We do not make assumptions on the confidence scores and expertise scores. We only need to assume that as J goes to infinity, $\bar{\Gamma}_J$ converges to a finite limit. If J is much larger than N (when $N/(N + J) \to 0$), we assume the smallest diagonal element Γ_0 is non-zero so that all component indeces receive sufficient input from the experts.

Table 4. Performance of composite index construction using cross-validation estimator.

	Sample Size $N = 400, J = 40$		
	PCA ($Q = 0$)	Q(CV)	Score ($Q = 10$)
MSE $\times 10^{-5}$	8.8756	7.5430	8.6774
RMSE $\times 10^{-3}$	8.8646	7.9328	8.2140
MSEr $\times 10^{-3}$	1.3537	1.0920	1.2119

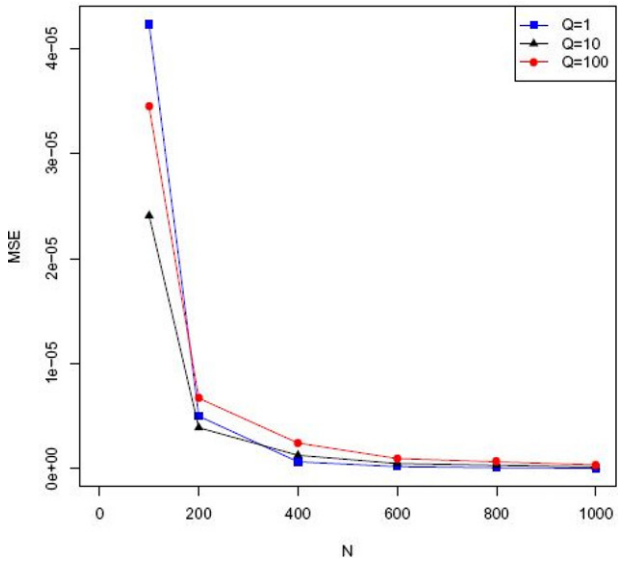

Figure 4. MSE against N.

Remark 7. Assumptions 1(*) is needed to handle the high-dimensional setting with $K \to \infty$, which is on the convergence rates of $\hat{\Sigma}$. Since it is not the focus of this article to consider the covariance matrix estimation, we list it as a high level condition. Such convergence rates often require structural assumptions on Σ_0, and have been extensively studied and are widely available in the literature; see, for example, Fan, Liao, and Liu (2016) and Vershynin (2018) and references therein, among many others. We also note that when K is fixed, Assumptions 1 guarantees that Assumption 1(*) is fulfilled with $\check{\Sigma}_N = \hat{\Sigma}_N$ and $\Delta_{1N} = N^{-1/2}$.

Table 5. Descriptive statistics of objective data.

Indices	Description	Mean	S.D.	Min	Max
x_{1i}	# Papers published in CSSCI and SCI / Total population of the region (10,000)	3.79	5.37	0.64	30.28
x_{2i}	# National achievements / Total population of the region (10,000)	119.61	306.48	0.00	1754.00
x_{3i}	# Invention Patents / Total population of the region (10,000)	9.07	15.70	1.30	85.00
x_{4i}	Technical Transfer Amounts (10,000 CNY) / Total population of the region (10,000)	1128.79	3039.22	0.00	16969.24
x_{5i}	International Technology Revenue (USD) / Gross Domestic Product (10,000 CNY)	4.20	10.63	0.02	52.42

Table 6. Correlation matrix.

Indices	$x_{1\cdot}$	$x_{2\cdot}$	$x_{3\cdot}$	$x_{4\cdot}$	$x_{5\cdot}$
$x_{1\cdot}$	1.0000	0.9698	0.9674	0.9763	0.7324
$x_{2\cdot}$	0.9698	1.0000	0.9521	0.9844	0.6354
$x_{3\cdot}$	0.9674	0.9521	1.0000	0.9386	0.7785
$x_{4\cdot}$	0.9763	0.9844	0.9386	1.0000	0.6150
$x_{5\cdot}$	0.7324	0.6354	0.7785	0.6150	1.0000

We first establish the convergence rate of $\hat{\boldsymbol{w}}$. For this result, we allow K to grow with N and J. Based on Remark 7, a fixed K is a special case of this scenario.

Theorem 1. Set the Q in Equation (5) as $Q = Q_N = \nu N\Delta_{1N}^2$, where $\nu > 0$ is a constant. Under Assumptions 1(*), 2, 3, 4(*), and 5, when $N \to \infty$ and $J \to \infty$, we have

$$||\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0|| = O_P\left(\min\{\Delta_{1N}, J^{-1/2}\}\right),$$

where $\hat{\boldsymbol{w}}_{N,J}$ is the maximizer of $g_{N,J}(\boldsymbol{w}, \delta)$ in Equation (5).

The proof is shown in the appendix.

Remark 8. As discussed earlier, the parameter Q balances the two sources of information. Theorem 1 requires $Q = Q_N = \nu N\Delta_{1N}^2$ to allow $\hat{\boldsymbol{w}}$ to follow the faster convergence rate. For an arbitrary Q_N, let $R_N = \sqrt{(JQ_N)/N}$, the convergence rate can be shown as

$$\min\left\{\Delta_{1N} \vee \left[(R_N \wedge 1)J^{-1/2}\right], \left[(1 \wedge R_N^{-1})\Delta_{1N}\right] \vee J^{-1/2}\right\}.$$

This is a slightly stronger result, but requires a more tedious proof.

When K is fixed, we can further have the central limit theorem for $\hat{\boldsymbol{w}} := \hat{\boldsymbol{w}}_{N,J}$.

Theorem 2. Under Assumptions 1 to 5, it holds that

$$(N+J)^{1/2}(\hat{\boldsymbol{w}} - \boldsymbol{w}_0) \Rightarrow N\left[0, \Lambda_0^{-1}(a\Omega_1 + (1-a)Q^2\delta_0^2\Omega_2)\Lambda_0^{-1}\right],$$

where $\Lambda_0 = a(\Sigma_0 - \lambda_1 \boldsymbol{I}) - (1-a)\delta_0^2 Q\mathcal{P}_2\Gamma_0 + \boldsymbol{w}_0\boldsymbol{w}_0'$, $\Omega_1 = \text{var}(\mathcal{P}_1\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{w}_0)$ and $\Omega_2 = \mathcal{P}_2\tilde{\Sigma}_s\mathcal{P}_2$, where $\mathcal{P}_1 = \boldsymbol{I} - \boldsymbol{w}_0\boldsymbol{w}_0'$ and $\mathcal{P}_2 = \boldsymbol{I} - (\boldsymbol{w}_0'\Gamma_0\boldsymbol{w}_0)^{-1}\Gamma_0\boldsymbol{w}_0\boldsymbol{w}_0'$. If $\boldsymbol{x}_i$ follows a normal distribution, then $\Omega_1 = \lambda_1(\Sigma_0 - \lambda_1\boldsymbol{w}_0\boldsymbol{w}_0')$.

The proof is shown in the appendix.

Remark 9. In the unbalanced cases (i.e., $N/(N+J)$ goes to 0 or 1), the combined estimator has the same asymptotic variance as that using the dominant source of information only. In the balanced case, the estimator is more efficient than using only one source.

Remark 10. The penalty coefficient Q should be chosen to minimize the trace of the asymptotic variance matrix. However, it is quite involved as Q appears in both Λ_0 and the middle term $(a\Omega_1 + (1-a)Q^2\Omega_2)$ in the asymptotic variance. In practice, we use cross validation procedure to choose the optimal Q.

4. Simulation Studies

In this section, we present some empirical studies to illustrate the performance of the proposed estimator $\hat{\boldsymbol{w}}_{N,J}$ under different N and J combinations. The impact of the penalty parameter Q and the performance of the cross-validation method are investigated as well.

For each of simulation, we assume $\boldsymbol{x}_i \sim N(0, \Sigma_0)$ where $\Sigma_0 = \boldsymbol{w}_0\boldsymbol{w}_0' + \sigma_\epsilon^2\boldsymbol{I}$, hence, it can be written as a factor model $\boldsymbol{x}_i = f_i\boldsymbol{w}_0 + \boldsymbol{\epsilon}_i$ where $f_i \sim N(0,1)$ and $\boldsymbol{\epsilon}_i \sim N(0, \sigma_\epsilon^2\boldsymbol{I})$. All f_i and $\boldsymbol{\epsilon}_i$ are iid and independent to each other. The expert scores $\boldsymbol{s}_j$ are generated according to the distribution of $\boldsymbol{s}$ described as follows. We first generate $\tilde{\boldsymbol{s}}$ through the spherical representation

$$\begin{aligned}\tilde{s}_1 &= \cos(e_1),\\ \tilde{s}_2 &= \sin(e_1)\cos(e_2), \ldots, \tilde{s}_K\\ &= \sin(e_1)\cdots\sin(e_{K-2})\sin(e_{K-1}),\end{aligned}$$

where the spherical coordinates e_k are IID $N(0, \sigma_s^2)$. Then, we choose a $K \times K$-dimensional orthogonal matrix $\boldsymbol{U}$ whose first column is $\boldsymbol{w}_0$, and generate $\boldsymbol{s}$ as $\boldsymbol{s} = \boldsymbol{U}\tilde{\boldsymbol{s}}$. It holds that $||\boldsymbol{s}|| = 1$, $\mathbb{E}(\boldsymbol{s}) = \delta_0\boldsymbol{w}_0$, where $\delta_0 = \mathbb{E}[\cos(e_1)] = \exp(-\sigma_s^2/2)$.

Exercise (i). In this experiment, we investigate the impact of the penalty parameter Q. Specifically, we use $K = 4$, $N = 400$, $J = 40$, $\gamma_{kj} = 1$ and $c_j = 1$ for $k = 1, \ldots, K, j = 1, \ldots, J$. We set $\boldsymbol{\theta}_0 = (\pi/6, \pi/4, \pi/3)$ with corresponding $\boldsymbol{w}_0 = (0.866025, 0.353553, 0.176777, 0.306186)$. The expert scores $\boldsymbol{s}_j$ are generated as described above. Here, we use $\sigma_\epsilon = 0.2, \sigma_s = 0.2$. We vary Q from 0 to 100 in the estimation.

Figure 2 shows the solution paths of the estimate $\hat{\boldsymbol{w}}$ as the penalty parameter Q changes. It is seen that the solution paths are continuous. The three horizontal lines mark the value of the PCA solution (using only the observed component indices), the estimates base on the experts score only, and the true value $\boldsymbol{w}_0$ in the factor model. The relationship between the estimated optimal combination weights and penalty parameter is clear seen. When $Q = 0$, the estimated combination weights are equal to the PCA estimates as the solution of the first term of the objective function (6). When Q increases, the estimated combination weights become closer to that of the experts. It is noted that the four solution paths cross the true combination weight line at different Q values and in the case of w_1 it does not cross at all. Table 1 lists some of the values.

Exercise (ii). We repeat the simulation in Exercise (i) 100 times, with sample sizes $(N, J) = (100, 10)$, $(400, 10)$, $(100, 40)$, and $(400, 40)$ to check to performance of the estimator $\hat{\boldsymbol{w}}_{N,J}$. We set $(\sigma_\epsilon, \sigma_s) = (0.2, 0.2)$. Again all c_j's and γ_{kj}'s are set to 1. Here we investigate the performance with Q ranging from 0 to 10.

Four performance criteria are used: mean squared error (MSE), root of mean squared error (RMSE), the relative of MSE (MSEr), and mean angle error (MAE). They are defined as

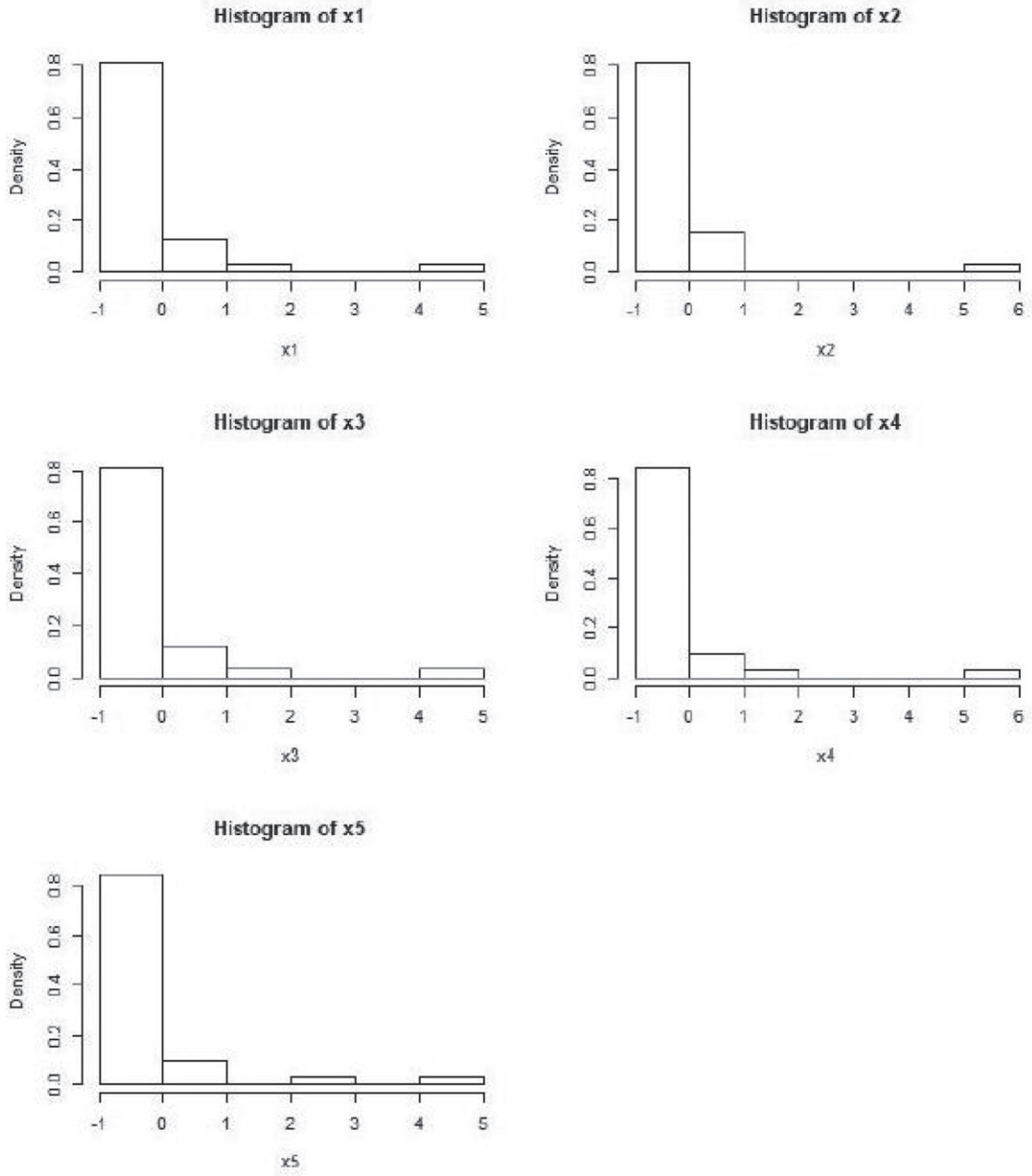

Figure 5. Distribution of component indices.

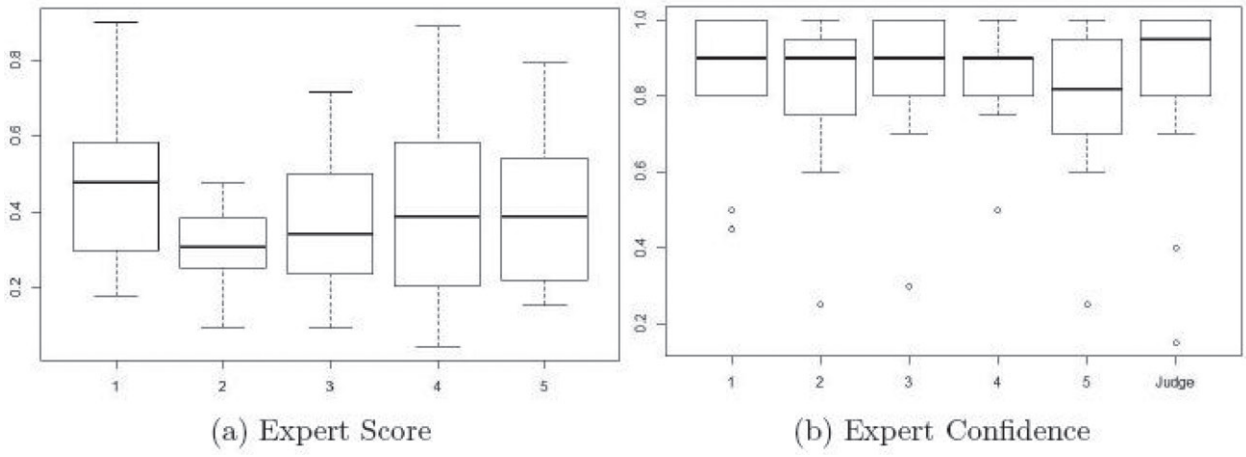

Figure 6. Expert opinion.

Table 7. Descriptive statistics of subjective data.

Scores	Mean	S.D.	Min	Max	Conf.	Mean	S.D.	Min	Max
s_{1j}	0.4812	0.2372	0.1782	0.9006	γ_{1j}	0.8423	0.1789	0.4500	1.0000
s_{2j}	0.3000	0.1208	0.0971	0.4735	γ_{2j}	0.8154	0.2125	0.2500	1.0000
s_{3j}	0.3677	0.1928	0.0937	0.7144	γ_{2j}	0.8354	0.1931	0.3000	1.0000
s_{4j}	0.4129	0.2665	0.0430	0.8909	γ_{2j}	0.8269	0.1666	0.5000	1.0000
s_{5j}	0.4142	0.1971	0.1537	0.7921	γ_{5j}	0.7954	0.2159	0.2500	1.0000
Expertise	Mean	S.D.	Min	Max					
c_j	0.8192	0.2650	0.1500	1.0000					

follows:

$$\text{MSE} = \frac{1}{L}\sum_{\ell=1}^{L} ||\hat{\boldsymbol{w}}^{(\ell)} - \boldsymbol{w}_0||^2,$$
$$\text{MAE} = \frac{1}{L}\sum_{\ell=1}^{L} ||\hat{\boldsymbol{w}}^{(\ell)} - \boldsymbol{w}_0||,$$
$$\text{MSEr} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{k=1}^{K}\left(\frac{\hat{w}_k^{(\ell)} - w_{0k}}{w_{0k}}\right)^2,$$
$$\text{RMSE} = \sum_{k=1}^{K}\left[\frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{w}_k^{(\ell)} - w_{0k}\right)^2\right]^{1/2},$$

where L is the number of replications. Here, RMSE gets the root mean square error of each component w_k first, then averages over k.

Figure 3 shows the four performance measures as functions of Q in the case of $(N, J) = (400, 40)$. All of them show a "U"-shape function, with minimum values corresponding to an optimal Q under different criteria. They demonstrate that by selecting an optimal Q, one can effectively combine both the observed data and the expert opinion for the construction of the composite index.

Table 2 shows the optimal Q and its corresponding performance under each criteria for the four sample size combinations. It is seen that, the optimal Q is larger when the number of experts J is larger, hence the combined experts' scores provides a more accurate estimate of the combination weights. When the sample size N is larger, the data provide more information, hence Q will be smaller.

Exercise (iii). In this experiment, we investigate the relationship between the optimal penalty parameter Q and the noise levels of the observations and expert opinions. Using $J = 10$, $K = 2$ and $N = 100$, we set $f_i \sim N(0, 1)$, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$, all $\boldsymbol{\gamma}$'s and c are set to 1. Let $\theta_0 = \pi/4$, or $\boldsymbol{w}_0 = (\sqrt{2}/2, \sqrt{2}/2)$, and we use different levels of σ_s to simulate expert scores $\boldsymbol{s}_j$.

Simulation is repeated 100 times to obtain MSE for each Q value and obtain the optimal Q $(0 \le Q \le 100)$ under each σ_ϵ and σ_s combination. Table 3 reports the optimal Q where σ_ϵ and σ_s are chosen to be the arithmetic sequence from 0.1 to 1 and from 0.05 to 0.50, separately.

There are some interesting observations. First, for a fixed σ_s, the optimal Q increases as σ_ϵ increases. It confirms our conclusion that the optimal estimator $(\hat{\boldsymbol{w}}_{N,J})$ should depend more on the expert opinions if the noise in the observed component indices is large. Second, for a fixed σ_ϵ, the optimal Q decreases as σ_s increases. This means that if the expert opinion is less reliable, the estimate $(\hat{\boldsymbol{w}}_{N,J})$ will have a higher relevance on the observed dataset. Hence, the optimal Q reflexes a balance between the noise level of the observed data and noise level of expert opinion information. If the σ_ϵ and σ_s are provided, then theoretical optimal Q can be chosen.

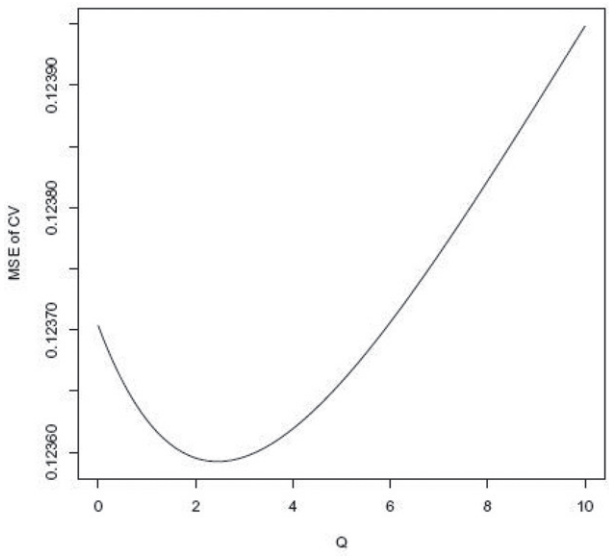

Figure 7. Cross-validation MSE.

Table 8. Estimation result.

Indices	w_1	w_2	w_3	w_4	w_5
Combined estimator	0.4497	0.4597	0.4440	0.4522	0.3707
Standard error	0.0255	0.0220	0.0142	0.0228	0.0397
PCA	0.4845	0.4747	0.4443	0.4621	0.3592
Standard error	0.2325	0.2301	0.2140	0.2223	0.1836
Weighted expert score	0.5829	0.3748	0.4314	0.3938	0.4227
Standard error	0.0749	0.0404	0.0590	0.0905	0.0634

Table 9. Composite Index.

Rank	Provinces	Index	Rank	Provinces	Index	Rank	Provinces	Index
1	Beijing	210.68	11	Sichuan	10.42	22	Hainan	5.15
2	Shanghai	88.64	12	Heilongjiang	10.18	23	Jiangxi	4.41
3	Tianjin	40.59	13	Shandong	8.91	24	Hebei	4.37
4	Jiangsu	23.35	14	Jilin	8.73	25	Ningxia	4.16
5	Shaanxi	22.37	15	Gansu	7.70	26	Shanxi	4.01
6	Guangdong	21.41	16	Qinhai	7.43	27	Yunnan	3.91
7	Zhejiang	17.93	17	Xinjiang	6.93	28	Guangxi	3.07
	Benchmark	16.71	18	Anhui	6.68	29	I. Mongolia	2.87
8	Hubei	16.58	19	Fujian	6.34	30	Guizhou	2.86
9	Liaoning	15.43	20	Hunan	6.18	31	Tibet	1.13
10	Chongqing	11.22	21	Henan	5.67			

Exercise (iv). In this exercise, we investigate the performance of the proposed cross-validation procedure for the determination of the optimal Q and its corresponding performance on the construction of the composite index. In this experiment, we assume $K = 4, N = 400, J = 40$, $f_i \sim N(0, 1)$ and $\boldsymbol{\epsilon}_i \sim N(0, 0.2^2)$. We set $\boldsymbol{\theta}_0 = (\pi/6, \pi/4, \pi/3)$ as Exercise (i). $\boldsymbol{s}_j$'s are generated using with $\sigma_s = 0.2$, $\gamma_{kj} = 1, c_j = 1$. We restrict Q in $[0, 10]$, and we use 10-fold cross-validation for N and 4-fold for J. Then we obtain CV(Q) in Equation (7), with $C = 1$.

Table 4 shows the performance of the estimator using the estimated optimal Q under cross-validation. The performance measures are obtained using 100 simulated datasets under each sample size setting. It is clearly seen that the combined construction with the optimal Q obtained from cross-validation outperforms that using data alone or using expert opinion alone.

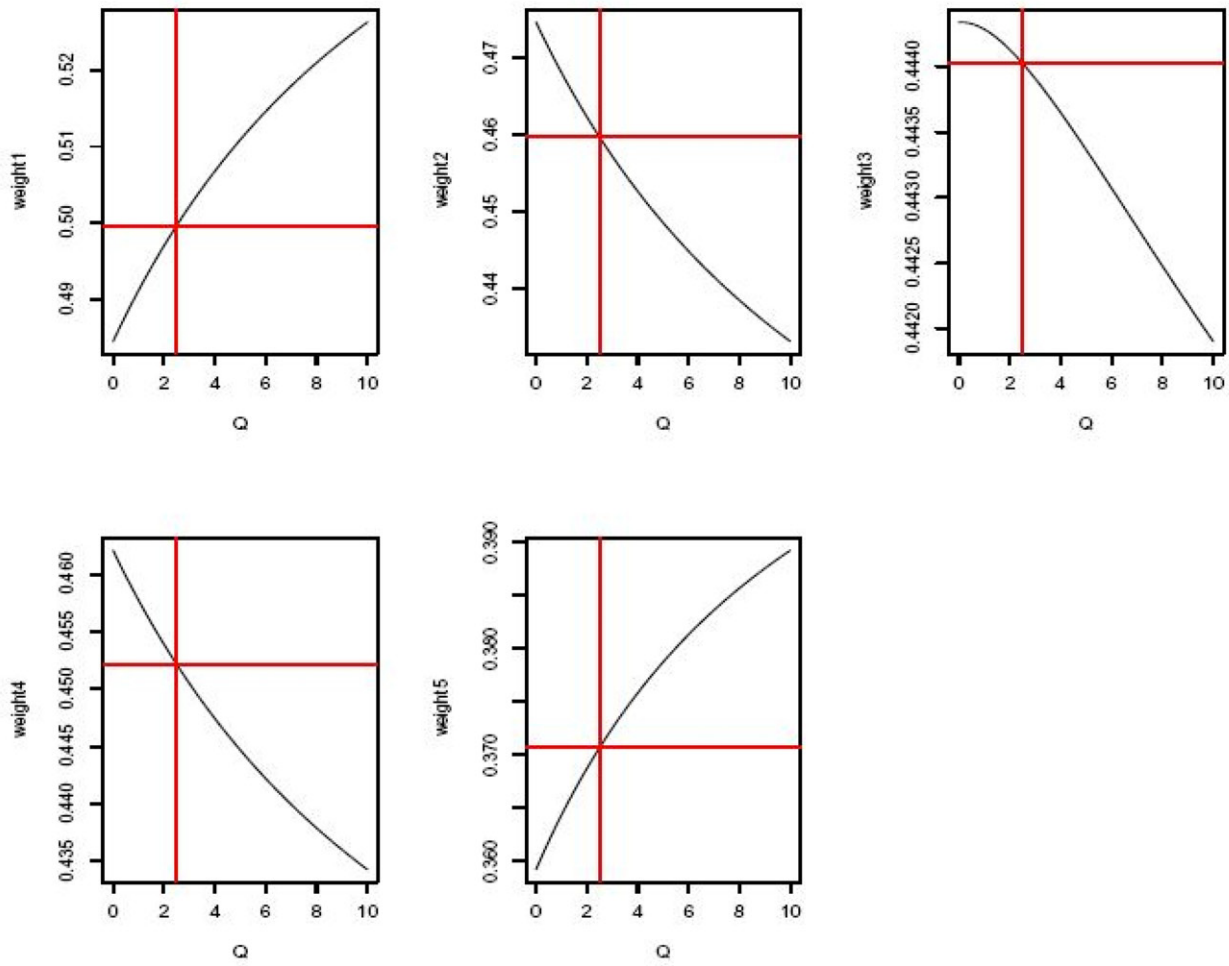

Figure 8. Solution path of real example.

Exercise (v). In this exercise, we investigate the convergence of $\hat{\boldsymbol{w}}_{N,J}$ under the high dimensional setup. We fix $J/N = 0.2$, let $K_N = 0.5N$ and generate $f_i \sim N(0,1)$, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$. We choose $\sigma_\epsilon^2 = \sigma_{\epsilon,N}^2 = 5/K_N$ so that the relative rank of Σ_{0N}, which is $\text{tr}(\Sigma_{0N})/||\Sigma_{0N}|| = 6/(1+5/K_N)$, stays roughly at a constant, and $||\hat{\Sigma}_N - \Sigma_{0N}|| = O_P(N^{-1/2})$ according to Koltchinskii and Lounici (2017), see also Chapter 9 of Vershynin (2018). The expert scores $\boldsymbol{s}_j$ are generated as *Exercise (i)* with all γ_{kj}'s and c_j setting to 1, $\boldsymbol{\theta}_0 = \pi * (1, 2, \ldots, K_N - 1)'$, and $\sigma_s = 0.2$. We plot the MSE of $\hat{\boldsymbol{w}}_{N,J}$ against N in Figure 4, for different choices of Q. The convergence of $\hat{\boldsymbol{w}}_{N,J}$ to $\boldsymbol{w}_0$ is clearly seen from the plot.

5. An Application

In this section, as an empirical application, we use the proposed method to construct a composite index for scientific and technological activity output of provinces and province-level municipalities in China. It is important for the policy makers to be able to evaluate the output of scientific and technological activities which in turn provide guidance for generating scientific and technological investment policies. In this example, we use five component indices (indicators) from National Scientific and Technological Progress Statistical Monitoring Database maintained by the Ministry of Science and Technology of China. The indicators include Number of Scientific Papers per capita (in 10,000 people), Number of National Scientific and Technological Achievements Awards per capita (in 10,000 people), Number of Invention Patents per capita (in 10,000 people), Technical Transfer Amounts (in 10,000 CNY) per capita (in 10,000 people), and International Technology Revenue (in USD) per Gross Domestic Product (in 10,000 CNY). We use data of year 2017. There are 31 provinces and province-level municipalities, excluding Taiwan, Hong Kong and Macao. Source of data is the 2017 National Scientific and Technological Progress Statistical Monitoring Report from Ministry of Science and Technology. Table 5 shows the detailed variable description and some descriptive statistics.

We standardize each of observed component index $\boldsymbol{x}_{k\cdot}$ to mean zero and standard deviation 1. Table 6 shows the correlation matrix of the five component indices. The correlations among the component indices are very close to one, except $\boldsymbol{x}_{5\cdot}$.

Figure 5 shows the distribution of the five indices. There are some obvious outliers. Among this small sample, Shanghai and Beijing are two very large outliers. They hold the two largest values in each of the five indicators. This is due to the fact that Shanghai and Beijing are the political, culture, science and technology, and business centers of China. These two outliers are also the source of the extreme high sample correlations among the component indices. We exclude these two sets of

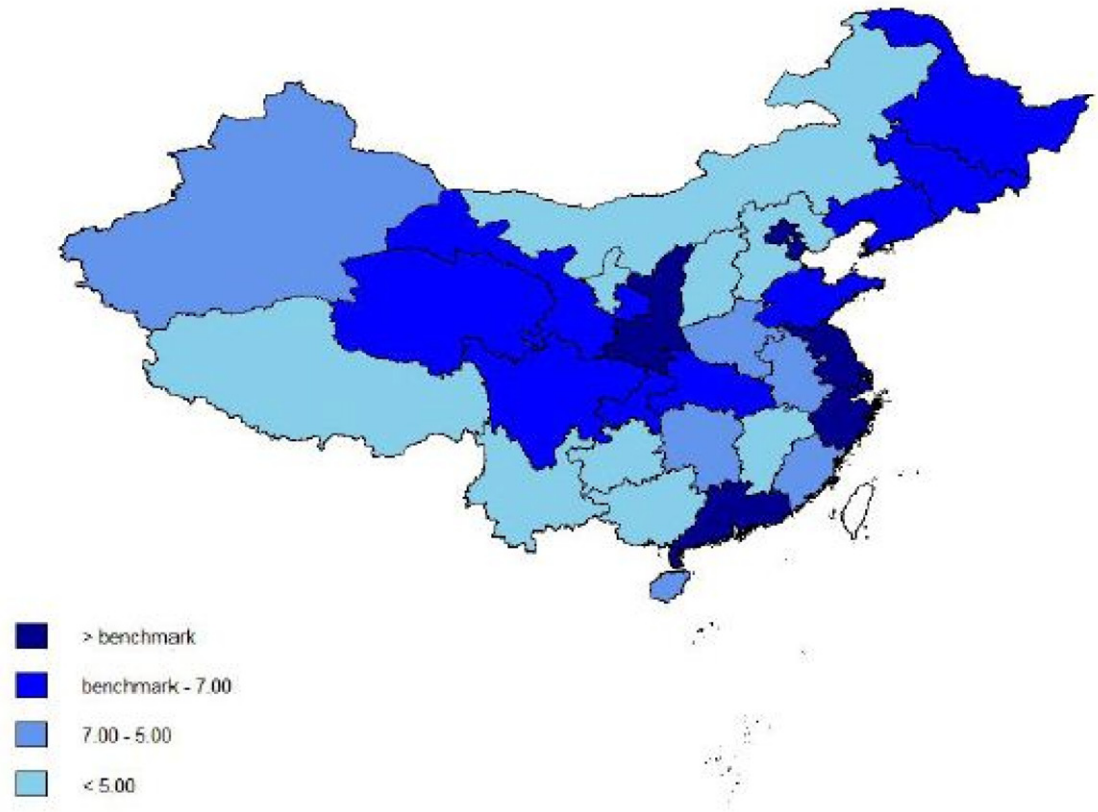

Figure 9. Composite Index for Science and Technology Output of China.

observations from the estimation of the combination weights for the composite index construction. Otherwise, their features will dominate the entire composition. We obtain the value of their composite index at the end base on the combination weights estimated using the data set without these two cities, for comparison purpose.

We surveyed 13 researchers who are experts on the issues related to science and technology development. Each expert gave scores to each of the five component indices and their corresponding confidence score. We also assessed their expertise levels ranging from 0.15 to 1. Figure 6 shows the boxplot of the expert scores and their confidence levels. The scores are standardized. Descriptive statistics of expert information is given in Table 7.

Since the sample size in this application is small, we use leave-one-out cross-validation instead of K-fold cross-validation in determining the optimal penalty parameter Q. Figure 7 shows $CV(Q)$ defined in (7). The estimated optimal Q is 2.5, when $CV(Q)$ is the smallest. Meanwhile, the estimated optimal δ is 0.8812.

Using the optimal Q, we estimate the combination weights, shown in Table 8. Due to the small sample size, we obtain bootstrap standard error (Efron and Tibshirani 1985) of the estimated combination weights, where both observed component induces and expert opinions are bootstrapped separately.

For combined estimation, we bootstrap the component indices and the export scores separately, in over the maintain the relative sample proportion. The combined estimates have relatively smaller bootstrap standard errors than the estimates using PCA or using expert opinion alone. Table 8 also shows the estimated combination weights using the observed data alone (PCA) and using the expert scores alone. For lower correlations between $x_{5.}$ and others, it is interesting to see that PCA gives a small $\hat{w}_5$ (0.3592), the combination weight for the international technology transfer, while the experts give a much larger value (0.4227). Combining both information, we assign 0.3707 to the indicator.

Figure 8 shows the solution path of the estimated combination weights as a function of Q. The vertical line indicates the optimal Q used and horizonal lines corresponding to the estimated combination weights.

The estimated combination weights are used to construct the composite index on science and technology output, including Beijing and Shanghai. We also use the national average of each component indicators to obtain the national index as a benchmark. Beijing has the highest science and technology output, since it has a large number of top universities and a large number of research institutes under Chinese Academy of Sciences. Shanghai ranks the second, due to its high concentration of major corporations and their R&D centers, as well as several major universities and research institutes. There are 7 provinces or province-level municipalities above the national benchmark: Beijing, Shanghai, Tianjin, Jiangsu, Shaanxi, Guangdong, and Zhejiang. Except Shannxi, these are the most industrial and developed regions of China. Figure 9 shows the constructed composite index for each provinces with their geographical locations. It is seen that some western provinces such as Shaanxi, Chongqing, Sichuan, Gansu, and Qinhai score high, although

traditionally their economic developments are slower than provinces on the east coast line. This is partially due to the recent strategic Western Development policy of the central government and the road-belt initiative (Liu and Dunford 2016; Démurger et al. 2002).

6. Conclusion

This article proposes a penalized optimization approach to incorporate expert opinion information with the principal component analysis of observed component indicators in a factor model framework for the construction of composite index using linear combination of the component indicators. The combination weights are determined by objective data and subjective expert opinion. The approach involves a penalty parameter Q that balances two sources of noises, one from the observations and the other from the expert opinions. It can be chosen through a data-driven cross-validation approach.

The proposed approach can be naturally and technically easily extended to construct multiple indices, similar to finding multiple factors or principle components. However, index construction often has specific target and interpretations — the reason that a group of experts would generally agree on the importance of each series. A second (and maybe orthogonal) index would be very difficult to interpret. In addition, it would be almost impossible to ask the experts to provide their weights on the second index that may or may not be orthogonal to the first one. Despite of the difficulty in defining and interpreting multiple indices, it is worth further exploration for practical uses.

Appendix

A.1. Proof of Theorem 1

Lemma 1. Under Assumption 2, $||\bar{\boldsymbol{s}}_J - \delta_0 \bar{\Gamma}_J \boldsymbol{w}_0|| = O_P(J^{-1/2})$.

Proof: Since $||\boldsymbol{s}_j|| = 1$ and all diagonal elements of Γ_j are less than or equal to 1,

$$\mathbb{E}||\Gamma_j \boldsymbol{s}_j - \delta_0 \Gamma_j \boldsymbol{w}_0||^2 \le \mathbb{E}||\Gamma_j \boldsymbol{s}_j||^2 \le 1.$$

Therefore, $\mathbb{E}||\bar{\boldsymbol{s}}_J - \delta_0 \bar{\Gamma}_J \boldsymbol{w}_0||^2 \le 1/J$, and the conclusion follows.

Proof of Theorem 1: Recall that $J = J_N$ is a constant depending implicitly on N. Set $\Delta_{2N} = J_N^{-1/2}$, and $\Delta_N = \min\{\Delta_{1N}, \Delta_{2N}\}$. To show that $||\hat{\boldsymbol{w}} - \boldsymbol{w}_0|| = O_P(\Delta_N)$, it suffices to prove that $||\hat{\boldsymbol{w}} - \hat{\boldsymbol{w}}_0|| = o_P(d_N \Delta_N)$ for any sequence $d_N \to \infty$. We shall prove that for any given diverging sequence $\{d_N\}$ and constant $\zeta > 0$,

$$\lim_{N\to\infty} P\left[\sup_{||\boldsymbol{w}-\boldsymbol{w}_0|| \ge \zeta d_N \Delta_N, \delta \in \mathbb{R}} g_{N,J}(\boldsymbol{w}, \delta) < g_{N,J}(\boldsymbol{w}_0, \delta_0)\right] = 1, \quad (8)$$

which implies $P\left(||\hat{\boldsymbol{w}} - \hat{\boldsymbol{w}}_0|| \ge \zeta d_N \Delta_N\right) = 0$, and then the conclusion of Theorem 1 follows.

Let

$$\begin{aligned} g_1(\boldsymbol{w}) &= \boldsymbol{w}' \check{\Sigma}_N \boldsymbol{w}, \\ g_2(\boldsymbol{w}, \delta) &= \delta^2 \boldsymbol{w}' \bar{\Gamma}_J \boldsymbol{w} - 2\delta \bar{\boldsymbol{s}}_J' \boldsymbol{w}. \end{aligned}$$

Recall that

$$g_{N,J}(\boldsymbol{w}, \delta) = a_{N,J} g_1(\boldsymbol{w}) - b_{N,J} Q_N g_2(\boldsymbol{w}, \delta).$$

Our strategy of proving (8) is to compare $g_{N,J}(\boldsymbol{w}, \delta)$ with $a_{N,J} g_1(\hat{\boldsymbol{w}}_1) - b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2)$ through

$$\begin{aligned} g_{N,J}(\boldsymbol{w}, \delta) = \; & a_{N,J} \left\{ g_1(\hat{\boldsymbol{w}}_1) - \left[g_1(\hat{\boldsymbol{w}}_1) - g_1(\boldsymbol{w}) \right] \right\} \\ & - b_{N,J} Q_N \left\{ g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - \left[g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - g_2(\boldsymbol{w}, \delta) \right] \right\}, \end{aligned}$$

where $\hat{\boldsymbol{w}}_1$ and $(\hat{\boldsymbol{w}}_2, \hat{\delta}_2)$ maximize $g_1(\boldsymbol{w})$ and $g_2(\boldsymbol{w}, \delta)$ respectively, whose precise definitions will be given in the sequel. Note that by definition

$$g_{N,J}(\boldsymbol{w}, \delta) \le a_{N,J} g_1(\hat{\boldsymbol{w}}_1) - b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2).$$

Let $\hat{\boldsymbol{w}}_1$ be the leading eigenvector of $\check{\Sigma}_N$. According to Wedin's $\sin(\Theta)$ Theorem (Wedin 1972),

$$||\hat{\boldsymbol{w}}_1 - \boldsymbol{w}_0|| \le \frac{\sqrt{2}||(\check{\Sigma}_N - \Sigma_{0N})\hat{\boldsymbol{w}}_1||}{||\check{\Sigma}_N|| - \lambda_{2N}} \le \frac{\sqrt{2}\,||\check{\Sigma}_N - \Sigma_{0N}||}{\lambda_{1N} - \lambda_{2N} - ||\check{\Sigma}_N - \Sigma_{0N}||},$$

which, together with Assumption 1(*), implies that $||\hat{\boldsymbol{w}}_1 - \boldsymbol{w}_0|| = O_P(\Delta_{1N})$. Write $\boldsymbol{w}_0 = \hat{\tau}\hat{\boldsymbol{w}}_1 + \sqrt{1-\hat{\tau}^2}\tilde{\boldsymbol{w}}_1$, where $0 \le \hat{\tau} \le 1$ and $\tilde{\boldsymbol{w}}_1 \perp \hat{\boldsymbol{w}}_1$. Note that $||\boldsymbol{w}_0 - \hat{\boldsymbol{w}}_1||^2 = 2(1-\hat{\tau})$. It holds that

$$\begin{aligned} g_1(\hat{\boldsymbol{w}}_1) - g_1(\boldsymbol{w}_0) &= \hat{\boldsymbol{w}}_1' \check{\Sigma}_N \hat{\boldsymbol{w}}_1 - \left(\hat{\tau}\hat{\boldsymbol{w}}_1 + \sqrt{1-\hat{\tau}^2}\tilde{\boldsymbol{w}}_1\right)' \\ &\qquad \check{\Sigma}_N \left(\hat{\tau}\hat{\boldsymbol{w}}_1 + \sqrt{1-\hat{\tau}^2}\tilde{\boldsymbol{w}}_1\right) \\ &= (1-\hat{\tau}^2)\hat{\boldsymbol{w}}_1' \check{\Sigma}_N \hat{\boldsymbol{w}}_1 - (1-\hat{\tau}^2)\tilde{\boldsymbol{w}}_1' \check{\Sigma}_N \tilde{\boldsymbol{w}}_1 \\ &\le 2(1-\hat{\tau})\hat{\lambda}_{1N} = \hat{\lambda}_{1N}\,||\boldsymbol{w}_0 - \hat{\boldsymbol{w}}_1||^2. \end{aligned}$$

By Assumption 1(*), there exists constants $\bar{\lambda} > \underline{\lambda} > 0$ such that $\lambda_{1N} \le \bar{\lambda}$ and $\lambda_{1N} - \lambda_{2N} \ge \underline{\lambda}$ when N is large enough. Consider the event A_{1N} on which

$$||\hat{\boldsymbol{w}}_1 - \boldsymbol{w}_0|| \le \sqrt{d_N/2}\Delta_{1N}, \quad \text{and} \quad \hat{\lambda}_{1N} < 2\bar{\lambda}, \quad \text{and} \quad |\hat{\lambda}_{1N} - \hat{\lambda}_{2N}| \ge \underline{\lambda}/2.$$

Then $P[A_{1N}] \to 1$, and on A_{1N},

$$g_1(\hat{\boldsymbol{w}}_1) - g_1(\boldsymbol{w}_0) \le \bar{\lambda} d_N \Delta_{1N}^2. \quad (9)$$

Recall that $\Delta_{2N} = J_N^{-1/2}$. Let $\hat{\delta}_2 = ||\bar{\Gamma}_J^{-1}\bar{\boldsymbol{s}}_J||$ and $\hat{\boldsymbol{w}}_2 = \bar{\Gamma}_J^{-1}\bar{\boldsymbol{s}}_J/\hat{\delta}_2$. By Assumption 4(*), there exists a constant $\underline{\gamma} > 0$ such that the minimum diagonal entry of $\bar{\Gamma}_J$ is larger than $\underline{\gamma}$ when N is large enough, and hence by Lemma 1,

$$||\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta_0\boldsymbol{w}_0|| \le ||\bar{\Gamma}_J^{-1}|| \cdot ||\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0|| = O_P(\Delta_{2N}).$$

Define the event $A_{2N} := \left[||\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta_0\boldsymbol{w}_0|| \le \sqrt{d_N}\Delta_{2N}\right]$, then $P[A_{2N}] \to 1$, and on A_{2N}, it holds that

$$\begin{aligned} g_2(\boldsymbol{w}_0, \delta_0) - g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) &= \left(\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta_0\boldsymbol{w}_0\right)\bar{\Gamma}_J\left(\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta_0\boldsymbol{w}_0\right) \\ &\le ||\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta_0\boldsymbol{w}_0||^2 \le d_N\Delta_{2N}^2. \end{aligned} \quad (10)$$

Combining (9) and (10), we know on the event $A_{1N} \cap A_{2N}$,

$$\begin{aligned} g_{N,J}(\boldsymbol{w}_0, \delta_0) \ge \; & a_{N,J} g_1(\hat{\boldsymbol{w}}_1) - b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) \\ & - d_N\left(a_{N,J}\bar{\lambda}\Delta_{1N}^2 + b_{N,J}Q_N\Delta_{2N}^2\right). \end{aligned} \quad (11)$$

We now consider $g_{N,J}(\boldsymbol{w}, \delta)$ with $||\boldsymbol{w}|| = 1$ and $||\boldsymbol{w} - \boldsymbol{w}_0|| \ge \zeta d_N \Delta_N$. Note that $||\boldsymbol{w}|| = 1$ and $||\boldsymbol{w} - \boldsymbol{w}_0|| \ge \zeta d_N \Delta_N$ implies

$$\inf_{\delta\in\mathbb{R}} ||\delta\boldsymbol{w} - \delta_0\boldsymbol{w}_0|| \ge \delta_0||\boldsymbol{w} - \boldsymbol{w}_0|| \ge \delta_0\zeta d_N\Delta_N.$$

Define two more events

$$\begin{aligned} A_{1N}' &= [||\boldsymbol{w} - \hat{\boldsymbol{w}}_1|| \ge \zeta d_N\Delta_N/2], \\ A_{2N}' &= [\inf_{\delta\in\mathbb{R}} ||\delta\boldsymbol{w} - \hat{\delta}_2\hat{\boldsymbol{w}}_2|| \ge \delta_0\zeta d_N\Delta_N/2]. \end{aligned}$$

When N is large enough, on $A_{1N} \cap A_{2N}$, at least one of A'_{1N} and A'_{2N} happens. More specifically, if $\Delta_{1N} \leq \Delta_{2N}$, then A'_{1N} happens, and thus (similar to (9)),

$$g_1(\hat{\boldsymbol{w}}_1) - g_1(\boldsymbol{w}) \geq 2(\hat{\lambda}_{1N} - \hat{\lambda}_{2N})||\boldsymbol{w} - \hat{\boldsymbol{w}}_1||^2 \geq \underline{\lambda}\zeta^2 d_N^2 \Delta_N^2/4.$$

Since $Q_N = \nu N\Delta_{1N}^2 = \nu(N\Delta_{1N}^2)/(J\Delta_{2N}^2)$, comparing with (11), we see that when N is large enough

$$\begin{aligned} g_{N,J}(\boldsymbol{w},\delta) &\leq a_{N,J} g_1(\hat{\boldsymbol{w}}_1) + b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - \underline{\lambda}\zeta^2 d_N^2 a_{N,J}\Delta_N^2/4 \\ &< a_{N,J} g_1(\hat{\boldsymbol{w}}_1) + b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - (\bar{\lambda} + \nu) d_N a_{N,J}\Delta_N^2 \\ &\leq g_{N,J}(\boldsymbol{w}_0, \delta_0). \end{aligned} \tag{12}$$

On the other hand, if $\Delta_{1N} \geq \Delta_{2N}$, then A'_{2N} happens, and

$$\begin{aligned} g_2(\boldsymbol{w},\delta) - g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) &= \left(\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta\boldsymbol{w}\right)\bar{\Gamma}_J\left(\hat{\delta}_2\hat{\boldsymbol{w}}_2 - \delta\boldsymbol{w}\right) \\ &\geq \underline{\gamma}\delta_0^2\zeta^2 d_N^2\Delta_N^2/4. \end{aligned}$$

Again, comparing with (11), when N is large enough,

$$\begin{aligned} g_{N,J}(\boldsymbol{w},\delta) &\leq a_{N,J} g_1(\hat{\boldsymbol{w}}_1) + b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - \underline{\gamma}\delta_0^2\zeta^2 d_N^2 b_{N,J}\Delta_N^2/4 \\ &< a_{N,J} g_1(\hat{\boldsymbol{w}}_1) + b_{N,J} Q_N g_2(\hat{\boldsymbol{w}}_2, \hat{\delta}_2) - (\bar{\lambda}/\nu + 1) d_N b_{N,J}\Delta_N^2 \\ &\leq g_{N,J}(\boldsymbol{w}_0, \delta_0). \end{aligned} \tag{13}$$

Combining (12) and (13), it holds that when N is large enough, on the event $A_{1N} \cap A_{2N}$, if $||\boldsymbol{w}|| = 1$ and $||\boldsymbol{w} - \boldsymbol{w}_0|| \geq \zeta d_N\Delta_N$, then $g_{N,J}(\boldsymbol{w},\delta) < g_{N,J}(\boldsymbol{w}_0,\delta_0)$, and the proof of (8) is complete, so is the proof of Theorem 1.

A.2. Proof of Theorem 2

Proof of Theorem 2: The Lagrangian form of optimizing $g_{N,J}(\boldsymbol{w},\delta)$ with $\boldsymbol{w}'\boldsymbol{w} = 1$ constraint is

$$a_{N,J}\boldsymbol{w}'\hat{\Sigma}_N\boldsymbol{w} - b_{N,J}Q\delta^2\boldsymbol{w}'\bar{\Gamma}_J\boldsymbol{w} + 2b_{N,J}Q\delta\bar{\boldsymbol{s}}_J'\boldsymbol{w} - a_{N,J}\lambda(||\boldsymbol{w}||_2^2 - 1).$$

Note that we insert $a_{N,J}$ in front of the Lagrangian term for ease of presentation. Its corresponding gradient condition is

$$a_{N,J}\hat{\Sigma}_N\hat{\boldsymbol{w}}_{N,J} - b_{N,J}Q\hat{\delta}^2\bar{\Gamma}_J\hat{\boldsymbol{w}}_{N,J} + b_{N,J}Q\hat{\delta}\bar{\boldsymbol{s}}_J - a_{N,J}\hat{\lambda}_{N,J}\hat{\boldsymbol{w}}_{N,J} = 0, \tag{14}$$

$$\hat{\delta}\hat{\boldsymbol{w}}_{N,J}'\bar{\Gamma}_J\hat{\boldsymbol{w}}_{N,J} - \bar{\boldsymbol{s}}_J'\hat{\boldsymbol{w}}_{N,J} = 0. \tag{15}$$

For notational simplicity, set $\tilde{N} = N + J$. By Theorem 1, we know under the assumptions of Theorem 2, it holds that

$$\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0 = O_P(\tilde{N}^{-1/2}), \quad \text{and} \quad \hat{\delta} - \delta = O_P(J^{-1/2}).$$

Therefore, the first term in (14) can be written as

$$\begin{aligned} a_{N,J}\hat{\Sigma}_N\hat{\boldsymbol{w}}_{N,J} &= a_{N,J}(\Sigma_0 + \hat{\Sigma}_N - \Sigma_0)(\boldsymbol{w}_0 + \hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \\ &= a_{N,J}\Sigma_0\boldsymbol{w}_0 + a_{N,J}(\hat{\Sigma}_N - \Sigma_0)\boldsymbol{w}_0 \\ &\quad + a_{N,J}\Sigma_0(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) + o_P(\tilde{N}^{-1/2}), \end{aligned} \tag{16}$$

due to the fact that $a_{N,J}(\Sigma_0 - \hat{\Sigma}_N) = O_P(a_{N,J}N^{-1/2}) = o_P(1)$. The sum of the second and third terms in (14) is

$$\begin{aligned} &- b_{N,J}Q\hat{\delta}^2\bar{\Gamma}_J\hat{\boldsymbol{w}}_{N,J} + b_{N,J}Q\hat{\delta}\bar{\boldsymbol{s}}_J \\ &= -b_{N,J}Q(\delta_0 + \hat{\delta} - \delta_0)^2\bar{\Gamma}_J(\boldsymbol{w}_0 + \hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \\ &\quad + b_{N,J}(\delta_0 + \hat{\delta} - \delta_0)Q(\delta_0\bar{\Gamma}_J\boldsymbol{w}_0 + \bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0) \\ &= -b_{N,J}Q\delta_0^2\bar{\Gamma}_J(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) + b_{N,J}Q\delta_0(\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0) \\ &\quad - b_{N,J}Q\delta_0(\hat{\delta} - \delta_0)\bar{\Gamma}_J\boldsymbol{w}_0 + o_P(\tilde{N}^{-1/2}), \end{aligned} \tag{17}$$

where we have used the fact

$$\begin{aligned} b_{N,J}Q(\hat{\delta} - \delta_0)(\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0) &= O_P(J/(N+J)\cdot J^{-1/2}\cdot J^{-1/2}) \\ &= o_P(\tilde{N}^{-1/2}) \end{aligned}$$

to get the second identity. The last term in (14) is treated under two scenarios: $a > 0$ and $a = 0$. First, if $a > 0$, (14) implies that $\hat{\lambda}_{N,J} - \lambda_1 = o_P(1)$, where λ_1 is the largest eigenvalue of Σ_0. Hence the last term in (14) is

$$\begin{aligned} a_{N,J}\hat{\lambda}_{N,J}\hat{\boldsymbol{w}}_{N,J} &= a_{N,J}(\lambda_1 + \hat{\lambda}_{N,J} - \lambda_1)(\boldsymbol{w}_0 + \hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \\ &= a_{N,J}\lambda_1\boldsymbol{w}_0 + a_{N,J}(\hat{\lambda}_{N,J} - \lambda_1)\boldsymbol{w}_0 \\ &\quad + a_{N,J}\lambda_1(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) + o_P(\tilde{N}^{-1/2}). \end{aligned} \tag{18}$$

Second, when $a = 0$, from (14), (16) and (17), we see that

$$a_{N,J}\hat{\lambda}_{N,J}\hat{\boldsymbol{w}}_{N,J} - a_{N,J}\lambda_1\boldsymbol{w}_0 = O_P(\tilde{N}^{-1/2}),$$

which implies that

$$\begin{aligned} a_{N,J}(\hat{\lambda}_{N,J} - \lambda_1)\hat{\boldsymbol{w}}_{N,J} &= a_{N,J}\lambda_1(\boldsymbol{w}_0 - \hat{\boldsymbol{w}}_{N,J}) + O_P(\tilde{N}^{-1/2}) \\ &= O_P(\tilde{N}^{-1/2}), \end{aligned}$$

and henceforth $a_{N,J}(\hat{\lambda}_{N,J} - \lambda_1) = O_P(\tilde{N}^{-1/2})$. Therefore, (18) will continue to hold when $a = 0$. From equation (15), using the facts $\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0 = O_P(\tilde{N}^{-1/2})$, $\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0 = O_P(J^{-1/2})$ and $\hat{\delta} - \delta_0 = O_P(J^{-1/2})$, we see that

$$\begin{aligned} (\hat{\delta} - \delta_0)\boldsymbol{w}_0'\bar{\Gamma}_J\boldsymbol{w}_0 + \delta_0\boldsymbol{w}_0'\bar{\Gamma}_J(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) - (\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0)'\boldsymbol{w}_0 \\ = o_P(\tilde{N}^{-1/2}), \end{aligned}$$

and it follows that

$$\hat{\delta} - \delta_0 = \frac{\boldsymbol{w}_0'(\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0) - \delta_0\boldsymbol{w}_0'\bar{\Gamma}_J(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0)}{\boldsymbol{w}_0'\bar{\Gamma}_J\boldsymbol{w}_0} + o_P(\tilde{N}^{-1/2}). \tag{19}$$

Plugging (19) into (17), and combining (16) and (18), we conclude that

$$\begin{aligned} &[a_{N,J}(\Sigma_0 - \lambda_1\boldsymbol{I}) - b_{N,J}Q\delta_0^2\bar{\mathcal{P}}_2\bar{\Gamma}_J](\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) - a_{N,J}(\hat{\lambda}_{N,J} - \lambda_1)\boldsymbol{w}_0 \\ &= -a_{N,J}(\hat{\Sigma}_N - \Sigma_0)\boldsymbol{w}_0 - b_{N,J}Q\delta_0\bar{\mathcal{P}}_2(\bar{\boldsymbol{s}}_J - \delta_0\bar{\Gamma}_J\boldsymbol{w}_0) + o_P(\tilde{N}^{-1/2}), \end{aligned} \tag{20}$$

where $\bar{\mathcal{P}}_2 = \boldsymbol{I} - (\boldsymbol{w}_0'\bar{\Gamma}_J\boldsymbol{w}_0)^{-1}\bar{\Gamma}_J\boldsymbol{w}_0\boldsymbol{w}_0'$. Multiplying both sides of (20) by the projection matrix $\mathcal{P}_1 = \boldsymbol{I} - \boldsymbol{w}_0\boldsymbol{w}_0'$ and using the facts $\mathcal{P}_1(\Sigma_0 - \lambda_1\boldsymbol{I}) = (\Sigma_0 - \lambda_1\boldsymbol{I})$and $\mathcal{P}_1\bar{\mathcal{P}}_2 = \bar{\mathcal{P}}_2$, we have

$$\begin{aligned} &\left[a_{N,J}(\Sigma_0 - \lambda_1\boldsymbol{I}) - b_{N,J}Q\delta_0^2\bar{\mathcal{P}}_2\bar{\Gamma}_J\right](\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \\ &= -a_{N,J}\mathcal{P}_1(\hat{\Sigma}_N - \Sigma_0)\boldsymbol{w}_0 - b_{N,J}Q\delta_0\bar{\mathcal{P}}_2(\bar{\boldsymbol{s}}_J - \bar{\Gamma}_J\boldsymbol{w}_0) + o_P(\tilde{N}^{-1/2}). \end{aligned}$$

Note that the matrix in front of $(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0)$ is singular. Since $||\hat{\boldsymbol{w}}_{N,J}||^2 = ||\boldsymbol{w}_0||^2 = 1$, it follows that $\boldsymbol{w}_0'(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) = (\boldsymbol{w}_0'\hat{\boldsymbol{w}}_{N,J} - 1) = -\frac{1}{2}||\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0||^2 = o_P(\tilde{N}^{-1/2})$. Therefore,

$$\begin{aligned} &\left[a_{N,J}(\Sigma_0 - \lambda_1\boldsymbol{I}) - b_{N,J}Q\delta_0^2\bar{\mathcal{P}}_2\bar{\Gamma}_J + \boldsymbol{w}_0\boldsymbol{w}_0'\right](\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \\ &= -a_{N,J}\mathcal{P}_1(\hat{\Sigma}_N - \Sigma_0)\boldsymbol{w}_0 - b_{N,J}Q\delta_0\bar{\mathcal{P}}_2(\bar{\boldsymbol{s}}_J - \bar{\Gamma}_J\boldsymbol{w}_0) + o_P(\tilde{N}^{-1/2}). \end{aligned} \tag{21}$$

Assumption 4 supposes $\bar{\Gamma}_J \to \Gamma_0$ and $J^{-1}\sum_{j=1}^J \Gamma_j\Sigma_s\Gamma_j' \to \tilde{\Sigma}_s$, which implies that $\bar{\mathcal{P}}_2 \to \mathcal{P}_2$. By the central limit theorem,

$$\begin{aligned} \sqrt{N}\mathcal{P}_1(\hat{\Sigma}_N - \Sigma_0)\boldsymbol{w}_0 &\Rightarrow N(0,\Omega_1), \\ \sqrt{J}\bar{\mathcal{P}}_2(\bar{\boldsymbol{s}}_J - \bar{\Gamma}_J\boldsymbol{w}_0) &\Rightarrow N(0,\Omega_2), \end{aligned}$$

where $\Omega_1 = \text{var}(\mathcal{P}_1 \boldsymbol{x}_i \boldsymbol{x}_i' \boldsymbol{w}_0)$ and $\Omega_2 = \mathcal{P}_2 \tilde{\Sigma}_s \mathcal{P}_2$. If $\boldsymbol{x}_i$ follows a normal distribution, then

$$\Omega_1 = \text{var}(\mathcal{P}_1 \boldsymbol{x}_i \boldsymbol{x}_i' \boldsymbol{w}_0) = \text{var}(\mathcal{P}_1 \boldsymbol{x}_i)\text{var}(\boldsymbol{x}_i' \boldsymbol{w}_0) = \lambda_1(\Sigma_0 - \lambda_1 \boldsymbol{w}_0 \boldsymbol{w}_0'),$$

due to the fact that $\mathcal{P}_1 \boldsymbol{x}_i$ and $\boldsymbol{x}_i' \boldsymbol{w}_0$ are both normal and $\mathbb{E}(\mathcal{P}_1 \boldsymbol{x}_i \boldsymbol{x}_i' \boldsymbol{w}_0) = 0$. Since the first two terms on the right-hand side of (21) are independent under Assumption 3, we have

$$\sqrt{N+J}(\hat{\boldsymbol{w}}_{N,J} - \boldsymbol{w}_0) \Rightarrow N\left[0, \Lambda_0^{-1}(a\Omega_1 + (1-a)Q^2\delta_0^2\Omega_2)\Lambda_0^{-1}\right],$$

where $\Lambda_0 = a(\Sigma_0 - \lambda_1 \boldsymbol{I}) - (1-a)Q\delta_0^2 \mathcal{P}_2 \Gamma_0 + \boldsymbol{w}_0 \boldsymbol{w}_0'$.

Funding

Financial supports in part by China Natural Science Foundation Grants 71773078, 71803134. We are also supported by the Innovative Research Team of Econometrics in Shanghai Academy of Social Sciences.

References

Aalianvari, A., Katibeh, H., and Sharifzadeh, M. (2012), "Application of Fuzzy Delphi AHP Method for the Estimation and Classification of Ghomrud Tunnel From Groundwater Flow Hazard," *Arabian Journal of Geosciences*, 5, 275–284. [67]

Al-Harbi, K. M. A. S. (2001), "Application of the AHP in Project Management," *International Journal of Project Management*, 19, 19–27. [67]

Alzate, C., and Suykens, J. A. (2010), "Multiway Spectral Clustering With Out-of-Sample Extensions Through Weighted Kernel PCA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 335–347. [67]

Chen, T. Y., and Li, C. H. (2010), "Determining Objective Weights With Intuitionistic Fuzzy Entropy Measures: A Comparative Analysis," *Information Sciences*, 180, 4207–4222. [67]

Cross, F. (1973), "The Behavior of Stock Prices on Fridays and Mondays," *Financial Analysts Journal*, 29, 67–69. [67]

Démurger, S., Sachs, J. D., Woo, W. T., Bao, S., Chang, G., and Mellinger, A. (2002), "Geography, Economic Policy, and Regional Development in China," *Asian Economic Papers*, 1, 146–197. [77]

Efron, B., and Tibshirani, R. (1985), "The Bootstrap Method for Assessing Statistical Accuracy," *Behaviormetrika*, 12, 1–35. [76]

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998), "Cluster Analysis and Display of Genome-Wide Expression Patterns," *Proceedings of the National Academy of Sciences*, 95, 14863–14868. [67]

Fan, J., Liao, Y., and Liu, H. (2016), "An Overview of the Estimation of Large Covariance and Precision Matrices," *The Econometrics Journal*, 19, C1–C32. [71]

Garivier, A. (2006), "Redundancy of the Context-Tree Weighting Method on Renewal and Markov Renewal Processes," *IEEE Transactions on Information Theory*, 52, 5579–5586. [68]

Hoskisson, R. E., Hitt, M. A., Johnson, R. A., and Moesel, D. D. (1993), "Construct Validity of an Objective (entropy) Categorical Measure of Diversification Strategy," *Strategic Management Journal*, 14, 215–235. [67]

Jiang, G., Liu, H., and Zhu, P. (1996), "The Maximum Variance Weighting Method Under the Constraint of Expert Opinion," *Statistical Research*, 6, 65–67. [68]

Jing, L., Ng, M. K., and Huang, J. Z. (2007), "An Entropy Weighting k-means Algorithm for Subspace Clustering of High-Dimensional Sparse Data," *IEEE Transactions on Knowledge & Data Engineering*, 19, 1026–1041. [67]

Karabel, J., and Astin, A. W. (1975), "Social Class, Academic Ability, and College Inequality," *Social Forces*, 53, 381–398. [67]

Kaur, A., and Lodhia, S. K. (2014), "The State of Disclosures on Stakeholder Engagement in Sustainability Reporting in Australian Local Councils," *Pacific Accounting Review*, 26, 54–74. [67]

Kawaller, I. G., Koch, P. D., and Koch, T. W. (1987), "The Temporal Price Relationship Between S&P 500 Futures and the S&P 500 Index," *The Journal of Finance*, 42, 1309–1329. [67]

Koltchinskii, V., and Lounici, K. (2017), "Concentration Inequalities and Moment Bounds for Sample Covariance Operators," *Bernoulli*, 23, 110–133. [75]

Kujawski, E. (2003), "Multi-Criteria Decision Analysis: Limitations, Pitfalls, and Practical Difficulties," *INCOSE International Symposium*, Vol 13, pp. 1169–1176. Wiley Online Library. [68]

Li, T., Zhang, H., Yuan, C., Liu, Z., and Fan, C. (2012), "A PCA-Based Method for Construction of Composite Sustainability Indicators," *The International Journal of Life Cycle Assessment*, 17, 593–603. [68]

Liu, W., and Dunford, M. (2016), "Inclusive Globalization: Unpacking China's Belt and Road Initiative," *Area Development and Policy*, 1, 323–340. [77]

Markowitz, H. (1956), "The Optimization of a Quadratic Function Subject to Linear Constraints," *Naval Research Logistics Quarterly*, 3, 111–133. [69]

Meade, L. M., and Presley, A. (2002), "R&D Project Selection Using the Analytic Network Process," *IEEE Transactions on Engineering Management*, 49, 59–66. [68]

Milligan, G. W. (1989), "A Validation Study of a Variable Weighting Algorithm for Cluster Analysis," *Journal of Classification*, 6, 53–71. [67]

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing. [68]

Rezaei, J. (2015), "Best-Worst Multi-Criteria Decision-Making Method," *Omega*, 53, 49–57. [68]

Roszkowska, E. (2013), *Rank Ordering Criteria Weighting Methods–A Comparative Overview*, Wydawnictwo Uniwersytetu w Białymstoku. [67]

Saaty, T. L. (2008), "Decision Making With the Analytic Hierarchy Process," *International Journal of Services Sciences*, 1, 83–98. [68]

Shemshadi, A., Shirazi, H., Toreihi, M., and Tarokh, M. J. (2011), "A Fuzzy VIKOR Method for Supplier Selection Based on Entropy Measure for Objective Weighting," *Expert Systems With Applications*, 38, 12160–12167. [67]

Stock, J. H., and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 4, 351–394. [67]

Tavoli, R., Kozegar, E., Shojafar, M., Soleimani, H., and Pooranian, Z. (2013), "Weighted PCA for Improving Document Image Retrieval System Based on Keyword Spotting Accuracy," in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 773–777. IEEE. [67]

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction With Applications in Data Science*, Vol. 47. Cambridge: Cambridge University Press. [71,75]

Wedin, P.-Å. (1972), "Perturbation Bounds in Connection With Singular Value Decomposition," *BIT Numerical Mathematics*, 12, 99–111. [77]

Whaley, R. E. (2008), Understanding VIX. Available at SSRN 1296743. [67]

Willems, F. M., Shtarkov, Y. M., and Tjalkens, T. J. (1995), "The Context-Tree Weighting Method: Basic Properties," *IEEE Transactions on Information Theory*, 41, 653–664. [68]

Yu, J., Yang, M. S., and Lee, E. S. (2011), "Sample-Weighted Clustering Methods," *Computers & Mathematics With Applications*, 62, 2200–2208. [67]

Zardari, N. H., Ahmed, K., Shirazi, S. M., and Yusop, Z. B. (2015), *Weighting Methods and Their Effects on Multi-Criteria Decision Making Model Outcomes in Water Resources Management*, Springer Publishing. [68]