mathematics

MDPI

Article

Permutation Variation and Alternative Hyper-Sphere Decomposition

Qingze Li [1,*,†] and Jianxin Pan [2,3,*,†]

[1] School of Mathematics and Statistics, Yunnan University, Kunming 650091, China
[2] Research Center for Mathematics, Beijing Normal University at Zhuhai, Zhuhai 519087, China
[3] Division of Science and Technology, United International College (BNU-HKBU), Zhuhai 519087, China
* Correspondence: liqingze@ynu.edu.cn (Q.L.); jianxinpan@uic.edu.cn (J.P.); Tel.: +86-1370-068-8818 (Q.L.)
† These authors contributed equally to this work.

Abstract: Current covariance modeling methods work well in longitudinal data analysis. In the analysis of data with no nature order, a common covariance modeling method would be inadequate. In this paper, a study is implemented to investigate the effects of permutations of data on the estimation of covariance matrix Σ. Based on the Hyper-sphere decomposition method (HPC), this study suggests that the change of data's permutation breaks the consistency of covariance estimation. An alternative Hyper-sphere decomposition method with permutation invariant is introduced later in this paper. The alternative method's consistency and asymptotic normality are studied when the observations follow a normal distribution. These results are tested using some example studies. Furthermore, a real data analysis is conducted for illustration purposes.

Keywords: order dependency; alternative Hyper-sphere decomposition; permutation invariant; unconstrained parameterization

Citation: Li, Q.; Pan, J. Permutation Variation and Alternative Hyper-Sphere Decomposition. *Mathematics* **2022**, *10*, 562. https://doi.org/10.3390/math10040562

Academic Editors: Min Wang, Haijun Gong, Liucang Wu, Songfeng Zheng and Jose Luis Vicente Villardon

Received: 18 December 2021
Accepted: 8 February 2022
Published: 11 February 2022

1. Introduction

The covariance matrix is a simple and popular method to describe the variation and correlation between random variables in multivariate statistics. A valid covariance matrix must be symmetric and semi-positive definite. In some areas, like social sciences, finance, economics and geology, people are more interested in studying the covariance matrices than mean models. Meanwhile, an appropriate working covariance matrix could increase the estimator's efficiency. Moreover, Pourahmadi (2013) [1] suggests a good estimate of the covariance matrix can lead to an accurate statistical inference and test results. Little and Rubin (2019) [2] show that an accurate estimate of covariance is essential in dealing with missing data problems. A sample covariance is usually used to estimate the population covariance matrix for its convenience. In some special cases, for example, high-dimensional situations where the number of repeated measurements n is less than that of unknown parameters in covariance matrix, using sample covariance matrix would end up with a biased estimator of mean model [3].

Pourahmadi (1999) [4] introduced the modified Cholesky decomposition method (MCD) into the estimation of covariance matrix by applying regression-based ideas into matrix decomposition methods. However, the interpretation of the relationship between model coefficients and correlation/variance of population Y is indirect. Pan and Pan (2017) [5] proposed the alternative Cholesky decomposition method (ACD) in 2017 that improves the correlation-interpretation problem of MCD by applying Cholesky decomposition on the correlation matrix R,

$$\Sigma = DRD = DTT'D,$$

where matrix D is a diagonal matrix with standard deviations σ of population Y, and T is a lower-triangular matrix with unit row vectors, $\vec{T}_{(i)}$ $i = 1, \ldots, p$, whose Euclidean norm is 1. The same model regression procedure as in MCD cannot be applied in ACD due to the extra unit-norm-restriction on row vectors in matrix T. Rebonato and Jäckel (2011) [6] projected

T in ACD into a unit Hyper-sphere coordinate system and proposed the Hyper-sphere decomposition method (HPC), in which elements t_{ij} in T are

$$t_{ij} = \begin{cases} 1 & i = j = 1 \\ \cos\phi_{ij} \prod_{k=1}^{j-1} \sin\phi_{ik} & j = 1, \ldots, p-1, i = 2, \ldots, p, \\ \prod_{k=1}^{j-1} \sin\phi_{ik} & j = p \end{cases} \quad (1)$$

where ϕ_{jk} are corresponding new angular coordinates of vector $\vec{T}_{(i)}$ in a unit hyper-sphere coordinate system. As a result we have

$$T = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ c_{21} & s_{21} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{p,1} & c_{p,2}s_{p,1} & c_{p,3}s_{p,2}s_{p,1} & \ldots & \prod_{k=1}^{p} s_{p,k} \end{pmatrix},$$

where c_{ij} and s_{ij} refer to $cos(\phi_{ij})$ and $sin(\phi_{ij})$. Comparing to ACD there is an additional matrix Φ, which is also a lower triangular with diagonal elements 0 and with the lower off-diagonal being angles $\phi_{ij} \in [0, \pi)$,

$$\Phi = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ \phi_{21} & 0 & 0 & \ldots & 0 & 0 \\ \phi_{31} & \phi_{32} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{p1} & \phi_{p2} & \phi_{p3} & \ldots & \phi_{p,p-1} & 0 \end{pmatrix}.$$

This method can guarantee the unit-row vectors in T for any given estimator matrix Φ; meanwhile, this decomposition has a geometry meaning for Φ [7]. In that sense, the regular model regression procedure can be applied in HPC just as in Pourahmadi (2000) [8]. To ensure $\phi_{jk} \in [0, \pi)$ we can do a further triangular transpose on the liner model:

$$tan(\phi_{jk} - \frac{\pi}{2}) = \omega'_{jk}\gamma \quad \phi \in [0, \pi), \quad \text{and} \quad \log\sigma_j^2 = h'_j\lambda, \quad (2)$$

where ω_{jk} and h_j are design matrices and vectors for angles ϕ_{jk} and variances σ_j^2, respectively. Meanwhile, γ and λ in Equation (2) are unknown correlation and variance components.

The current popular methods like HPC work well in longitudinal data analysis in which there is a natural order of the sample data. While in analyses like geometric and causal data there is no natural permutation of data, these methods generate estimates that depend on the permutation of sample data. The purpose of this article is to illustrate the permutation variance of covariance estimator of HPC. After that, we redefine the translation between T and Φ, then propose an alternative Hyper-sphere decomposition (AHPC) method that inherits most of the advantages of HPC while improving the permutation variance. Meanwhile, the AHPC has a more straightforward geometrical interpretation.

2. Interpretations of Hyper-Sphere Decomposition Method

Rapisarda, Brigo and Mercurio (2007) [7] explained HPC from the aspect of Jacobi rotation. In a p-dimensional right-hand system, axis $e_1, e_2, \ldots, e_p$ are selected as follows:

1. Set the direction of $\vec{T}_{(1)}$ as the first axis e_1.
2. In the plane containing vectors $\vec{T}_{(1)}$ and $\vec{T}_{(2)}$, make the e_2 vertical to e_1 in a right-hand system.
3. e_3 is perpendicular to the previous axis, e_1 and e_2, and on the same side of their cross product $e_1 \times e_2$.
4. After the signed j axis, the next axis e_{j+1} is defined as vertical to all previous axes and on the same direction of $e_1 \times e_2 \cdots \times e_j$, while $j + 1 \leq p$.

Cross product $e_j \times e_k = ||e_j|| \cdot ||e_k|| sin\theta \cdot \vec{n}$ determines a direction $\vec{n}$ perpendicular to the plane containing e_j and e_k given by the right-hand rule.

With all axes having been set up, the row vectors $\vec{T}_{(j)} = (t_{j1}, \ldots, t_{jj}, 0, \ldots, 0)'$ shows us the coordinate of $\vec{T}_{(j)}$ in this Cartesian system. According to [7], the relation between angles ϕ_{ij} and coordinates of $\vec{T}_{(k)}$s is (Figure 1a):

1. We begin with $\vec{T}_{(1)}$. Next we turn $\vec{T}_{(1)}$ counter-clockwise in hyper-plane PL_{12} with angle $\phi_{2,1}$. Then we have $\vec{T}_{(2)} = G(1,2;\phi_{2,1})\vec{T}_{(1)}$. The hyper-plane PL_{12} is built with e_1 and e_2.
2. Turn $\vec{T}_{(1)}$ in PL_{12} with angle $\phi_{3,1}$; then do another rotation in PL_{23} with angle $\phi_{3,2}$. We then have $\vec{T}_{(3)} = G(2,3;\phi_{3,2})G(1,2;\phi_{3,1})\vec{T}_{(1)}$.
3. $\vec{T}_{(k)}$ is built up with $k-1$ times turn as: $\vec{T}_{(k)} = G(k-1,k;\phi_{k,k-1}) \ldots G(1,2;\phi_{k,1})\vec{T}_{(1)}$.

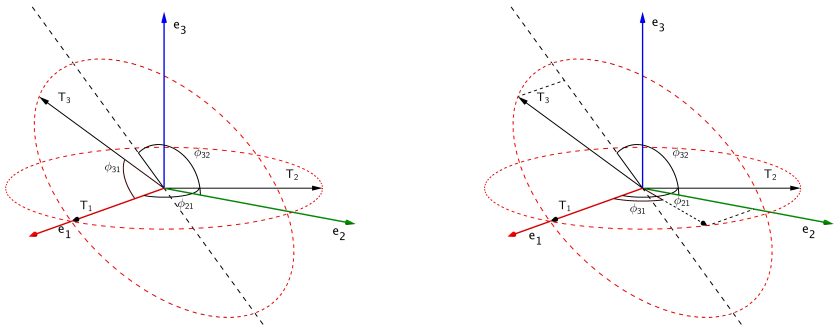
In a p-dimensional space, $G(j,k;\theta)$ is a Jacobi rotation matrix in hyper-plane $PL_{j,k}$ with counter-clockwise angle θ. Besides the geometric explanation from [7], we also propose a new interpretation of the correlation between matrices T and Φ in the HPC method, regarding angles between two hyper-planes in a row space of T (Figure 1b):

1. Lower off-diagonal entries in the first column of Φ, ϕ_{i1}s, are angles between the first and ith row vectors in T.
2. The lower off-diagonal entries in second column of Φ contains angles, ϕ_{i2}s, between plane HP_{12}, which is defined by the first and second row vectors in T, and HP_{1i} $i > 2$, which is defined by the first and ith row vectors in T.
3. After that, elements ϕ_{jk} in kth column of matrix Φ, $j > k \geq 3$, are angles between two hyper-planes $HP_{12\ldots(k-1)k}$ and $HP_{12\ldots(k-1)j}$.

The $HP_{12\ldots k}$'s are built with the first, second and until the kth row vectors. The angle between hyper-planes is well known in the mathematics literature, see, e.g., [Chep. 3] of Murty [9]. The angle ϕ_{jk}'s between two hyper-planes, $HP_{12\ldots(k-1)k}$ and $HP_{12\ldots(k-1)j}$ can be calculated recursively via

$$\theta_{jk} = \begin{cases} arcos(t_{jk}) & j \geq 2, k = 1 \\ arcos(\dfrac{t_{jk}}{t_{j(k-1)} tan(\theta_{j(k-1)})}) & j > 2, k \geq 2 \end{cases}, \quad (3)$$

According to [7], when θ_{jk}'s and ϕ_{jk}'s are in $[0, \pi]$, the translation in Equation (3) is unique, $\theta_{jk} = \phi_{jk}$.



(**a**) (**b**)

Figure 1. 3-dimensional illustration of HPC. (**a**) Interpretation of HPC method based on the idea of angles between hyper-plane. (**b**) Interpretation of HPC method based on Jacobi rotation.

3. Order-Dependence of Hyper-Sphere Decomposition Method

Based on both explanations of the HPC method above, we show that the HPC is order-dependent from two aspects in this section.

Setting R as a correlation matrix between random sample $Y = (Y_1, \ldots, Y_p)'$, and after a standard Cholesky decomposition $R = TT'$, where $T = \left(\vec{T}_{(1)}, \ldots, \vec{T}_{(p)}\right)'$, based on the definition of HPC decomposition, the correlation ρ_{jk} in R is actually the *cosine* value of an angle between two unit vectors $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$, $\rho_{jk} = \vec{T}'_{(j)}\vec{T}_{(k)} = cos < \vec{T}_{(j)}, \vec{T}_{(k)} >$. So eventually we say that these unit vectors $\vec{T}_{(j)}$ in lower-triangular matrix T represent the whole correlation information of corresponding random sample Y_j. In that sense, the order of row vectors in T has almost the same representation of the order of Y_j in correlation matrix R. For example, when $p = 6$, under the order $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$, we have R and the lower triangular T matrix in (4) where, for example, t_{42} is the coordinate on axis e_2 for vector $\vec{T}_{(4)}$, which corresponds to Y_4. Then we will have a Φ matrix as the left one in (5) where in the matrix Φ, from left to right, ϕ_{jk} is the kth Jacobi rotation angle in hyper-plane $PL_{k,k+1}$, which is built with e_k and e_{k+1}. For example, ϕ_{42} is the second Jacobi rotation angle in PL_{23}.

$$R = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{array} \begin{array}{c} \begin{array}{cccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 \end{array} \\ \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} & \rho_{16} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} & \rho_{35} & \rho_{36} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 & \rho_{45} & \rho_{46} \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1 & \rho_{56} \\ \rho_{61} & \rho_{62} & \rho_{63} & \rho_{64} & \rho_{65} & 1 \end{pmatrix} \end{array}, T = \begin{array}{c} \\ \vec{T}_{(1)} \\ \vec{T}_{(2)} \\ \vec{T}_{(3)} \\ \vec{T}_{(4)} \\ \vec{T}_{(5)} \\ \vec{T}_{(6)} \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ t_{21} & t_{22} & 0 & 0 & 0 & 0 \\ t_{31} & t_{32} & t_{33} & 0 & 0 & 0 \\ t_{41} & t_{42} & t_{43} & t_{44} & 0 & 0 \\ t_{51} & t_{52} & t_{53} & t_{54} & t_{55} & 0 \\ t_{61} & t_{62} & t_{63} & t_{64} & t_{65} & t_{66} \end{pmatrix} \end{array}. \quad (4)$$

From the aspect of angles between two hyper-planes, we have the table for matrix Φ as in the right one in (5) where j stands for row number. In this table, for example, ϕ_{53} is the angle between $HP_{1,2,3}$ and $HP_{1,2,5}$. Hyper-plane $HP_{1,2,5}$ is built with the first, second and fifth row vectors in T.

$$\begin{array}{c} \\ \vec{T}_{(1)} \\ \vec{T}_{(2)} \\ \vec{T}_{(3)} \\ \vec{T}_{(4)} \\ \vec{T}_{(5)} \\ \vec{T}_{(6)} \end{array} \begin{array}{c} \begin{array}{cccccc} e_1e_2 & e_2e_3 & e_3e_4 & e_4e_5 & e_5e_6 & \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_{21} & 0 & 0 & 0 & 0 & 0 \\ \phi_{31} & \phi_{32} & 0 & 0 & 0 & 0 \\ \phi_{41} & \phi_{42} & \phi_{43} & 0 & 0 & 0 \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & 0 & 0 \\ \phi_{61} & \phi_{62} & \phi_{63} & \phi_{64} & \phi_{65} & 0 \end{pmatrix} \end{array}, \begin{array}{c} \\ \vec{T}_{(1)} \\ \vec{T}_{(2)} \\ \vec{T}_{(3)} \\ \vec{T}_{(4)} \\ \vec{T}_{(5)} \\ \vec{T}_{(6)} \end{array} \begin{array}{c} \begin{array}{ccccc} 1|j & 12|1j & \ldots & 12345|1234j & \end{array} \\ \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ \phi_{21} & 0 & \ldots & 0 & 0 \\ \phi_{31} & \phi_{32} & \ldots & 0 & 0 \\ \phi_{41} & \phi_{42} & \ldots & 0 & 0 \\ \phi_{51} & \phi_{52} & \ldots & 0 & 0 \\ \phi_{61} & \phi_{62} & \ldots & \phi_{65} & 0 \end{pmatrix} \end{array}. \quad (5)$$

Eventually we have the same equation between ρ and ϕ as in [10]:

$$\rho_{jk} = c_{jk} \prod_{l=1}^{k-1} s_{jl} s_{kl} + \sum_{l=1}^{k-1} \left[c_{jl} c_{kl} \prod_{t=1}^{l-1} s_{jt} s_{kt} \right], \quad (6)$$

where s_{jk} and c_{jk} stand for $sin(\phi_{jk})$ and $cos(\phi_{jk})$, respectively.

If we change the order of Y_js in correlation matrix R, simultaneously we change the corresponding row vectors, $\vec{T}_{(j)}$s, in lower-triangular matrix T. While the elements in new matrix T^* may change after changing order, the relationship between $\vec{T}_{(j)}$ and Y_j remains the same. As illustrated before, elements in matrix T are coordinates. When the order of Y_js is different, we set up a new set of axes. That leads us to have a new set of coordinates for $\vec{T}_{(j)}$s. If we change the order into $Y_1, Y_2, Y_5, Y_4, Y_3, Y_6$, then the new correlation matrix R^* can be obtained by swapping elements in R and doing the Cholesky decomposition on R^* we get T^* as in (7).

$$R^* = \begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_5 \\ Y_4 \\ Y_3 \\ Y_6 \end{array} \begin{array}{c} \begin{array}{cccccc} Y_1 & Y_2 & Y_5 & Y_4 & Y_3 & Y_6 \end{array} \\ \begin{pmatrix} 1 & \rho_{12} & \rho_{15} & \rho_{14} & \rho_{13} & \rho_{16} \\ \rho_{21} & 1 & \rho_{25} & \rho_{24} & \rho_{23} & \rho_{26} \\ \rho_{51} & \rho_{52} & 1 & \rho_{54} & \rho_{53} & \rho_{56} \\ \rho_{41} & \rho_{42} & \rho_{45} & 1 & \rho_{43} & \rho_{46} \\ \rho_{31} & \rho_{32} & \rho_{35} & \rho_{34} & 1 & \rho_{36} \\ \rho_{61} & \rho_{62} & \rho_{65} & \rho_{64} & \rho_{63} & 1 \end{pmatrix} \end{array}, T^* = \begin{array}{c} \\ \vec{T}_{(1)} \\ \vec{T}_{(2)} \\ \vec{T}_{(5)} \\ \vec{T}_{(4)} \\ \vec{T}_{(3)} \\ \vec{T}_{(6)} \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & e_2 & e_3^* & e_4^* & e_5^* & e_6^* \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ t_{21} & t_{22} & 0 & 0 & 0 & 0 \\ t_{31} & t_{32} & * & 0 & 0 & 0 \\ t_{41} & t_{42} & * & * & 0 & 0 \\ t_{51} & t_{52} & * & * & * & 0 \\ t_{61} & t_{62} & * & * & * & * \end{pmatrix} \end{array}. \quad (7)$$

Since we changed the order from position Y_3, simultaneously a new set of axes e_3^*, e_4^*, e_5^* and e_6^* are selected. As a result we will have new value for the coordinate on e_3^*, e_4^* and e_5^*, represented as $*$s. Meanwhile, the corresponding Φ^* matrix will be like in (8).

$$\begin{array}{c} \\ \tilde{T}_1' \\ \tilde{T}_2' \\ \tilde{T}_5' \\ \tilde{T}_4' \\ \tilde{T}_3' \\ \tilde{T}_6' \end{array} \begin{array}{c} \begin{array}{cccccc} e_1e_2 & e_2e_3^* & e_3^*e_4^* & e_4^*e_5^* & e_5^*e_6^* & \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \phi_{21} & 0 & 0 & 0 & 0 & 0 \\ \phi_{51} & \phi_{52} & 0 & 0 & 0 & 0 \\ \phi_{41} & \phi_{42} & * & 0 & 0 & 0 \\ \phi_{31} & \phi_{32} & \phi_{53} & * & 0 & 0 \\ \phi_{61} & \phi_{62} & * & * & * & 0 \end{pmatrix} \end{array}, \begin{array}{c} \\ \vec{T}_{(1)} \\ \vec{T}_{(2)} \\ \vec{T}_{(5)} \\ \vec{T}_{(4)} \\ \vec{T}_{(3)} \\ \vec{T}_{(6)} \end{array} \begin{array}{c} \begin{array}{ccccc} 1|j & 12|1j & 123|12j & \dots & \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \phi_{21} & 0 & 0 & \dots & 0 \\ \phi_{51} & \phi_{52} & 0 & \dots & 0 \\ \phi_{41} & \phi_{42} & * & \dots & 0 \\ \phi_{31} & \phi_{32} & \phi_{53} & \dots & 0 \\ \phi_{61} & \phi_{62} & * & \dots & 0 \end{pmatrix} \end{array}. \quad (8)$$

From the aspect of Jacobi rotation, element ϕ_{jk} in Φ is a rotating angle in plane $PL_{k,k+1}$ built by axes e_k and e_{k+1}. Because the definition of axis e_js depend on the order of $\vec{T}_{(j)}$, which is again also the order of Y_j, ϕ_{jk} still depends on the order of Y_j. We can observe that $cos < \vec{T}_{(1)}, \vec{T}_{(k)} >= \vec{T}'_{(1)}\vec{T}_{(k)} = (1, \dots, 0)(c_{k1}, \dots .0)' = c_{k1} = f(\phi_{k1})$, and $cos < \vec{T}_{(2)}, \vec{T}_{(k)} >= \vec{T}'_{(2)} \cdot \vec{T}_{(k)} = (c_{21}, \dots, 0)(c_{k1}, \dots, 0)' = c_{21}c_{k1} + s_{21}c_{k2}s_{k1} = f(\phi_{21}, \phi_{k1}, \phi_{k2})$, which shows that ρ_{1k} and ρ_{2k} only relay on ϕ_{k1}, ϕ_{k2} and ϕ_{21}. This explains the reason why ϕ_{k1} and ϕ_{k2} remain the same, respectively. Since the third vector becomes $\vec{T}_{(5)}$, angles on the column 3 all will have a new value, except ϕ_{53}, for obvious reasons, the angle between $HP^*_{1,2,3}$ and $HP^*_{1,2,5}$ equals that between $HP_{1,2,5}$ and $HP_{1,2,3}$.

We summarize this dependency in Table 1. Table 1 indicates how the value of entries in Φ depend on the order of row vectors in matrix T.

After proving the permutation variation of the values in the Φ matrix, we consider how this would affect our modeling process. If the HPC method were permutation invariant, the model assumption should always hold no matter how we change the order of data Y. However, a different order gives different values of Φ, proven above. Then, under the model (2), with the same covariates ω^* and h^* in which the corresponding values have been rearranged according to the new order, HPC would have a different estimator of $\hat{\gamma}$ and $\hat{\lambda}$. In some cases, even the whole model assumption would be wrong under the new permutation of sample data.

Table 1. Dependence between entries in Φ and row vectors in T. For example, 1, 2, 3, p means element ϕ_{p3}, at row p column 3 in Φ depends on the first, second, third and pth row vectors in T.

	col1	**col2**	**col3**	**…**	**colp-1**	**colp**
row1	0	0	0	…	0	0
row2	1, 2	0	0	…	0	0
row3	1, 3	1, 2, 3	0	…	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
rowp	1, p	1, 2, p	1, 2, 3, p	…	1, 2, 3, …, p − 1, p	0

Due to the definition of translation between matrices T and Φ, according to our study, the data order affects Φ element-wise. Furthermore, there is no way of transforming a new angle matrix under the new data order back to its previous one only by rearranging elements. In other words, a change of the order will give a different set of values for Φ. The regression model on Φ elements would have different parameter estimates under varied data order. Furthermore, even the model assumption may vary according to the data order. In that sense, we say the current HPC method is order dependent. As the conclusion, under both interpretations, the HPC method is permutation variate both in terms of the Φ matrix and models for variance components.

4. Alternative Hyper-Sphere Decomposition Method

We notice that the order dependency of the HPC method is caused by the definition of the translation between two coordinate systems, the Cartesian coordinate system for matrix T and a unique angular coordinate system for Φ. We now propose a new definition of matrix Φ inspired by spherical parameterization [11], which would improve the relationship between matrices R and Φ. Then eventually the newly proposed alternative Hyper-sphere decomposition (AHPC) would be permutation invariant.

4.1. The AHPC Model

The AHPC method differs from HPC in a new definition on Φ, the lower triangular matrix with diagonal elements equal to 0. Instead of being angles between hyper-planes in HPC, here we define our elements in Φ as angles between row vectors $\vec{T}_{(j)}$s, as illustrated in Figure 2,

$$\phi_{jk} = < \vec{T}_{(j)}, \vec{T}_{(k)} > \qquad 1 < k < j \leq p, \tag{9}$$

where $\phi_{jk} \in [0, \pi)$ to ensure the uniqueness.

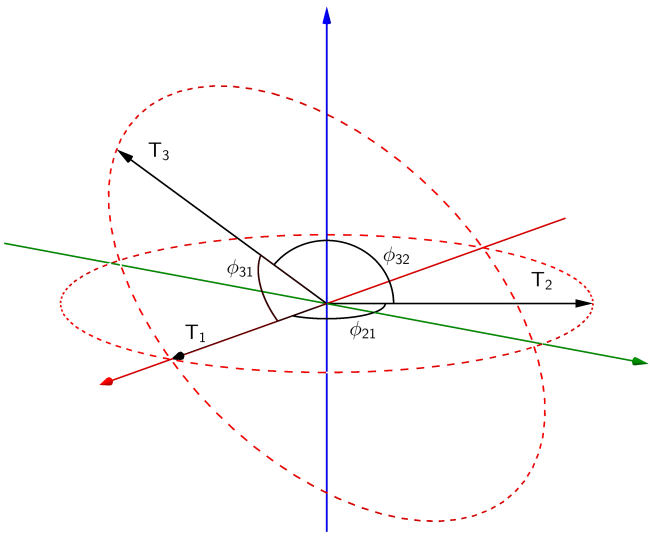


Figure 2. A geometric representation of AHPC when $p = 3$. The ϕ_{jk}s are angles between $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$, respectively.

For dimension $p \geq 3$, there is a three-dimensional restriction. Any pairs of vectors T_i and T_j must be in a two-dimensional plan, the angle $\phi_{ij} \in [0, \pi)$ between $\vec{T}_{(i)}$ and $\vec{T}_{(j)}$, to guarantee the uniqueness. Furthermore, for angles between three vectors, there are only two proper situations. The first one is $\phi_{jk} = \phi_{ji} + \phi_{ik}$ when vector $\vec{T}_{(i)}$, $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$ are in the same plane, and $\vec{T}_{(i)}$ is sitting between $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$. The second scenario is $\phi_{jk} < \phi_{ji} + \phi_{ik}$ where three vectors are not in the same plane. Consequently, for any three vectors $\vec{T}_{(i)}$, $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$, angles between each pair of vectors must satisfy that the sum of

any two angles be not less than the third one and the difference between any two angles be not greater than the third one,

$$\phi_{jk} \leq \phi_{ji} + \phi_{ik} \text{ and } \phi_{jk} \geq \phi_{ji} - \phi_{ik}, \quad k < i < j, \tag{10}$$

where $\phi_{ji} = \phi_{ij}$. Thus for the restriction on angles between vectors in three-dimensional spaces, the model assumption must satisfy the constraint $\phi_{jk} \leq \phi_{ij} + \phi_{ik}$. This is not only due to the geometric fact but also is sufficient for positive semi-definiteness of correlation matrix R. When $\phi_{jk} = \phi_{ji} + \phi_{ik}$, T_i, T_j and T_k must be on the same plane, which suggests that there must exist constants a and b such that $T_i = aT_j + bT_k$. As a result, in the ith row of correlation matrix R,

$$\rho_{ih} = T_i' T_h = (aT_j + bT_k)' T_h = aT_j' T_h + bT_k' T_h \quad 1 \leq j \leq p,$$

which indicates the ith, jth and kth rows are linearly dependent; then the corresponding correlation matrix R would be positive semi-definite, $det(R) = 0$. On the other hand, if $\phi_{jk} < \phi_{ji} + \phi_{ik}$, R would be positive definite.

We model the variance components by:

$$\log \sigma_j^2 = h_j' \lambda, \qquad \phi_{jk} = g(\omega_{jk}), \tag{11}$$

where $g(\omega_{jk})$ should be a monotonic decreasing function for stationary data whose correlation only depends on absolute distances between covariates $\omega_{jk} = ||Location_j - Location_k||$. For angles between three vectors $\vec{T}_{(i)}$, $\vec{T}_{(j)}$ and $\vec{T}_{(i)}$, the distance ω_{jk} is less than $\omega_{ji} + \omega_{ik}$; then $\phi_{jk} = g(\omega_{jk}) \leq g(\omega_{ji}) + g(\omega_{ik}) = \phi_{ji} + \phi_{ik}$. Another option is optimization with nonlinear programming (NLP). We can use NLP methods [12], with nonlinear inequality constraints,

$$\begin{array}{ll} g(\omega_{jk}) - g(\omega_{ji}) - g(\omega_{ik}) + s_{jk}^2 = 0, & j < i < k, \\ |g(\omega_{ji}) - g(\omega_{ik})| - g(\omega_{jk}) + s_{jk}^2 = 0, & j < i < k, \end{array} \tag{12}$$

to calculate the maximum likelihood estimator. The s_{jk} are slack parameters in the Karush–Kuhn–Tucker (KKT) condition for NLP to translate inequality constraints into equality ones. Libraries like *fmincon* in MATLAB and program *LINGO* could deal with this NLP problem as well. However, this is more of a mathematical issue than a statistical problem. So no details on NLP will be offered and studied in this paper.

Under the new definition of Φ, we have a direct relation between R and Φ:

$$\rho_{jk} = \rho_{kj} = cos(\phi_{jk}), \tag{13}$$

where ϕ_{jk} is in $[0, \pi]$. This is the key to guarantee permutation invariance. Comparing with Equation (6), ρ_{jk} in AHPC only depends on one corresponding component ϕ_{jk} instead of depending on a series of $\phi_{jk}s$ in a triangular area over $j+1$ rows in HPC. The translations between T and Φ are

$$cos(\phi_{jk}) = c_{jk} = cos < \vec{T}_{(j)}, \vec{T}_{(k)} > = \vec{T}_{(j)}' \vec{T}_{(k)} = \sum_{l=1}^{k} t_{jl} t_{kl} \qquad j > k, \tag{14}$$

and

$$t_{jk} = \begin{cases} c_{j1} & j > k = 1 \\ \left[c_{jk} - (\sum_{l=1}^{k-1} t_{jl} t_{kl}) \right] / (t_{kk}) & j > k > 1 \\ 1 & j = k = 1' \\ \left[1 - (\sum_{l=1}^{k-1} t_{jl}^2) \right]^{1/2} & j = k > 1 \end{cases} \tag{15}$$

where t_{jk}s are elements in T.

When changing R into R^* by re-ordering Y in matrix form, actually we apply a transformation matrix A on R such that:

$$ARA^T = R^*,$$

in which R and R^* are non-negative definite symmetric and A is a square matrix. Under this setting, finding A solves the unique form of the algebraic Riccati equation

$$AC_1^T C_1 A = R^* + F^T A + AF, \tag{16}$$

where $C_1^T C_1 = R$, which means C_1 is the Cholesky decomposition of R and setting $F = 0$. Consequently, there must exist a transformation matrix A such that $ARA^T = R^*$.

Meanwhile, based on the definition of the permutation matrix, P, when changing the order of rows in matrix R from $1, 2, 3, 4$ to R^* with new order $4, 1, 3, 2$,

$$PRP^T = R^*, \tag{17}$$

where P is a permutation matrix. Thus, permutation matrix P is a solution of transformation matrix A. In that sense, after Cholesky decomposition $R = LL^T$ and $R^* = L^* L^{*T}$, the equation below is true:

$$PRP^T = PLL^T P^T = L^* L^{*T} = R^*. \tag{18}$$

Eventually, we have $PL = L^*$, noting that we need to simplify PL into a lower triangular matrix. We turn PL into a lower-triangular matrix, equivalent to selecting a new set of the coordinate axis. Furthermore, that is what makes L^* differ from L.

In the alternative hyper-sphere decomposition method, there is a direct link between correlation ρ_{jk} and its corresponding coefficient ϕ_{jk} in function (13). Only in the proof below, we use the symmetric $\Phi_f = \Phi + \Phi^T$ instead of the lower triangular one. Thus, we can have:

$$PRP^T = Pcos(\Phi_f)P^T = cos(P\Phi_f P^T) = cos(\Phi_f^*) = R^*,$$

which also means $P\Phi_f P^T = \Phi_f^*$, where Φ_f^* is the angle matrix under the new order. For example, changing $1, 2, 3, 4$ into $4, 1, 2, 3$:

$$P\Phi_f P^T = \begin{pmatrix} 0 & \phi_{41} & \phi_{43} & \phi_{42} \\ \phi_{14} & 0 & \phi_{13} & \phi_{12} \\ \phi_{34} & \phi_{31} & 0 & \phi_{32} \\ \phi_{24} & \phi_{21} & \phi_{23} & 0 \end{pmatrix} = \Phi_f^*.$$

We can observe that entries in Φ_f and Φ_f^* are identical with locations rearranged. This guarantees the Φ matrix in AHPC is permutation invariant.

Make model assumption on variance components as in (2) with covariate matrices Ω and H. After changing into the new order, correspondingly we rearrange our new Ω_f^* as $\Omega_f^* = P\Omega_f P^T$. As proven before, the value of ϕ_{jk} remains the same. Its corresponding ω_{jk} and h_j under the model assumption also remains the same. We can see that the model assumption and the estimators $\hat{\lambda}$ & $\hat{\gamma}$ will remain the same after any change of the order.

As shown above, from both the aspect of entries in Φ and the model for variance components, the AHPC method is permutation invariant.

4.2. Estimation Method for AHPC

We use the two algorithms to get the maximum likelihood estimator of mean parameter β and covariance components γ and λ jointly. Giving response $Y_{p\times n} = (Y_1, Y_2, \ldots, Y_n)'$ that with p measurements and each measurement has n observations, under the linear model assumption, we have:

$$Y - X\beta \sim MVN(0, \Sigma), \tag{19}$$

where X is a design matrix, and β is an unknown parameter vector. In AHPC without loss of generality, we can set up a joint regression model for mean and covariance components as:

$$\mu_j = x_j'\beta, \qquad \log\sigma_j^2 = h_j'\lambda, \qquad \phi_{jk} = \omega_{jk}'\gamma, \tag{20}$$

where μ_j and σ_j ($j = 1, 2, \ldots, p$) are, respectively, the mean and standard deviation for the jth measurement Y_j. Moreover, ϕ_{jk}s $\in [0, \pi]$ ($1 < k < j \leq p$) are angles in matrix Φ. Dimensions for β, λ and γ are $q_\beta, q_\lambda, q_\gamma$, respectively. Ignoring the constant $-(np/2)\log(2\pi)$, the minus twice log-likelihood function has the following representation:

$$\begin{array}{rl} -2l & = n\log|DRD| + \sum_{i=1}^{n} r_i' D^{-1} R^{-1} D^{-1} r_i \\ & = n\log|DRD| + \sum_{i=1}^{n} \xi_i' \xi_i, \end{array} \tag{21}$$

where r_is are residuals $r_i = Y_i - \mu$ and $\xi_i = T^{-1}D^{-1}r_i$. In AHPC we have

$$\log|DRD| = \log|DTT'D| = 2\log|D| + 2\log|T| = \sum_{j=1}^{p}(\log\sigma_j^2 + 2\log t_{jj}),$$

where t_{jj}s are diagonal elements in T. We define Δ as a $p \times p$ matrix, and $\Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m})$ means the elements of this matrix are $\frac{\partial \rho_{jk}}{\partial \gamma_m}$s ($1 \leq j, k \leq p$; $m = 1, 2, \ldots, q_\gamma$). Since $\Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m})$ is symmetric as R, $\Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m}) = \Delta'(\frac{\partial \rho_{jk}}{\partial \gamma_m})$. Taking the first derivative of the log-likelihood function above with respect to β, λ and γ separately, we can obtain the following score functions:

$$\begin{array}{rl} -2\frac{\partial l}{\partial \beta} & = 2\sum_{i=1}^{n} X_i' \Sigma^{-1} r_i, \\ -2\frac{\partial l}{\partial \lambda} & = n\sum_{l=1}^{p} h_l - \sum_{i=1}^{n}\sum_{l=1}^{p} \xi_{ij} \sum_{k=1}^{j} a_{lk} r_k \frac{h_k}{\sigma_k}, \\ -2\frac{\partial l}{\partial \gamma_m} & = ntr(R^{-1}\Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m})) - \sum_{i=1}^{n} tr(r_i' D^{-1} R^{-1} \Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m}) R^{-1} D^{-1} r_i), \end{array} \tag{22}$$

where ξ_{jk} ($1 \leq j \leq n$; $1 \leq k \leq p$) is the kth element in ξ_j, a_{lk}s are the (l, k)th elements in T^{-1}. Based on model assumption (20), $\frac{\partial \rho_{jk}}{\partial \gamma_m} = -sin(\phi_{jk})\omega_{jkm}$, where ω_{jkm} is the mth element in vector ω_{jk}; and $\partial l/\partial \gamma = (\partial l/\partial \gamma_1, \partial l/\partial \gamma_2, \ldots, \partial l/\partial \gamma_{q_\gamma})'$. Set score functions above or equal to zero to get the estimator of parameters. In general, the score function does not have explicit solutions. Numerical optimization procedures such as the Newton–Raphson algorithm can be used here.

By calculating the second derivatives of the log-likelihood function with respect to β, λ and γ we can have the Fisher-information matrix I,

$$I = \begin{pmatrix} I_{\beta\beta'} & I_{\beta\lambda'} & I_{\beta\gamma'} \\ I_{\beta\lambda'}' & I_{\lambda\lambda'} & I_{\lambda\gamma'} \\ I_{\beta\gamma'}' & I_{\lambda\gamma'}' & I_{\gamma\gamma'} \end{pmatrix}. \tag{23}$$

in which,

$$I_{\beta\beta'} = \frac{\partial^2 l}{\partial \beta \partial \beta'} = nX'\Sigma^{-1}X. \tag{24}$$

The expected Fisher-information matrix of λ is as follows:

$$I_{\lambda\lambda'} = -E(\frac{\partial^2 l}{\partial \lambda \partial \lambda'}) = \frac{n}{4}\sum_{l=1}^{p}\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} a_{lk_1}a_{lk_2}\rho_{k_1k_2}(h_{k_1}h_{k_1}' + h_{k_2}h_{k_1}'). \tag{25}$$

Similarly to how we deal with the score function for γ,

$$-I_{\gamma\gamma'} = \begin{pmatrix} \frac{\partial^2 l}{\partial\gamma_1\partial\gamma_1'} & \frac{\partial^2 l}{\partial\gamma_1\partial\gamma_2'} & \cdots & \frac{\partial^2 l}{\partial\gamma_1\partial\gamma_{q_\gamma}'} \\ \frac{\partial^2 l}{\partial\gamma_2\partial\gamma_1'} & \frac{\partial^2 l}{\partial\gamma_2\partial\gamma_2'} & \cdots & \frac{\partial^2 l}{\partial\gamma_2\partial\gamma_{q_\gamma}'} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial\gamma_{q_\gamma}\partial\gamma_1'} & \frac{\partial^2 l}{\partial\gamma_{q_\gamma}\partial\gamma_2'} & \cdots & \frac{\partial^2 l}{\partial\gamma_{q_\gamma}\partial\gamma_{q_\gamma}'} \end{pmatrix} \tag{26}$$

We have:

$$\begin{aligned} -2\frac{\partial^2 l}{\partial\gamma_{m_1}\partial\gamma_{m_2}'} = \; & n \times tr\left(R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_{m_2}})R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_{m_1}})\right) \\ & + \textstyle\sum_{i=1}^n tr\left(r_i'D^{-1}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_{m_1}} + \frac{\partial\rho_{jk}}{\partial\gamma_{m_2}})R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_{m_1}} + \frac{\partial\rho_{jk}}{\partial\gamma_{m_2}})R^{-1}D^{-1}r_i\right) \end{aligned} \tag{27}$$

We use the same method to find the cross section parts with γ:

$$\begin{aligned} I_{\gamma_m\beta} = I_{\beta\gamma_m} \; & = -\textstyle\sum_{i=1}^n tr\left(\frac{\partial r_i'}{\partial\beta}D^{-1}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}r_i\right) \\ & = \textstyle\sum_{i=1}^n tr\left(X'D^{-1}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}r_i\right). \end{aligned} \tag{28}$$

and

$$\begin{aligned} I_{\gamma_m\lambda} = I_{\lambda\gamma_m} \; & = -\textstyle\sum_{i=1}^n tr\left(r_i'\frac{\partial D^{-1}}{\partial\lambda}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}r_i\right) \\ & = 2\textstyle\sum_{i=1}^n tr\left(r_i'D^{-1}HR^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}r_i\right) \end{aligned}, \tag{29}$$

where $H = (h_1, h_2, \ldots, h_p)'$. We can show that the expectations of off block diagonal matrices $I_{\beta\lambda}$ and $I_{\beta\gamma}$ are equal to $\mathbf{0}$. Detail of derivation of score functions and expected Fisher-information matrix can be found in Supplementary Materials.

Giving the expected Fisher-information matrix in the form:

$$I = \begin{pmatrix} I_{\beta\beta'} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{\lambda\lambda'} & I_{\lambda\gamma'} \\ \mathbf{0} & I_{\lambda\gamma'}' & I_{\gamma\gamma'} \end{pmatrix},$$

The Fisher-scoring algorithm can be applied here. This algorithm conveniently gives us the asymptotic covariance matrix of the estimator $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\gamma}$, the inverse Fisher-information matrix evaluated at the **MLE**s.

We propose the simple regression algorithm based on the direct relationship between ρ_{jk} and ϕ_{jk} as in function (13). We can use sample variance and the correlation matrix to estimate covariance coefficients λ and γ, even if the sample covariance matrix is singular. This approach cannot be applied to the HPC method for the requirement that the correlation matrix be positive-definite for the necessity of Cholesky decomposition. Otherwise there is no direct way to translate R into Φ.

This estimation method is more computationally efficient compared to the Newton–Raphson algorithm. Since we use **LSE** and **GWLSE**, there is no optimal searching problem, and also the grand optimality is guaranteed. As we prove before, parameter β in the mean model is independent of the covariance coefficients λ and γ, so there is no problem separating the estimation procedures as above.

Algorithm 1: Newton–Raphson Algorithm

Input: Initial value β_0, λ_0 and γ_0

repeat

Create the starting covariance matrix Σ_0 using λ_0 and γ_0

Using the update functions:

$$\beta_1 = \beta_0 + I^{-1}_{\beta\beta'}(\partial l/\partial\beta)_{\beta=\beta_0}, \tag{30}$$

and

$$\begin{pmatrix} \lambda_1 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} \lambda_0 \\ \gamma_0 \end{pmatrix} + \begin{pmatrix} I_{\lambda\lambda'} & I_{\lambda\gamma'} \\ I'_{\lambda\gamma'} & I_{\gamma\gamma'} \end{pmatrix}^{-1} \begin{pmatrix} \partial l/\partial\lambda \\ \partial l/\partial\gamma \end{pmatrix}_{\gamma=\gamma_0,\lambda=\lambda_0} \tag{31}$$

to update parameter estimate β_1, λ_1 and γ_1

Set new initial value as β_1, λ_1 and γ_1

until *Convergence for all parameters* β, λ *and* γ

Algorithm 2: Simple Regression Algorithm

Calculate sample mean $\bar{Y}$ and initial residual $r_i = Y_i - \bar{Y}$

repeat

Calculate sample standard diviation matrix $D = diag(\sigma_i)$ and correlation R with residual r_i

Transform R into Φ by $\phi_{jk} = arccos(\rho_{jk})$. And use D and Φ to find the **LSE** $\hat{\lambda}_{(1)}$ and $\hat{\gamma}_{(1)}$

Create estimator $\hat{\Sigma}$ with $\hat{\lambda}_{(1)}$ and $\hat{\gamma}_{(1)}$. And find the Generalized Weighted Least Squares Estimator (**GWLSE**) of $\hat{\beta}$

Use $\hat{\beta}$ to update $\bar{Y}$ and r_i

until *Convergence of* $\hat{\beta}$, $\hat{\lambda}$ *and* $\hat{\gamma}$ *is reached*

5. Asymptotic Property

In this section, consistency and asymptotic normality of AHPC estimators under standard situation $p < n$ are proven, similar to the proof in [10], under some regularity conditions.

Theorem 1. *Under the regularity conditions when sample size n goes to* ∞*, the maximum likelihood estimator* $\hat{\beta}$, $\hat{\lambda}$ *and* $\hat{\gamma}$ *is consistent with the true value,* $(\hat{\beta}', \hat{\lambda}'.\hat{\gamma}')' \xrightarrow{a.s} (\beta'_0, \lambda'_0, \gamma'_0)'$.

We already have the score functions for β, λ and γ. Based on the consistency of **MLE**s $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\gamma}$ in Theorem 1 , we then have:

$$E(\frac{\partial l}{\partial\beta})_{\hat{\beta}} = 0, \qquad E(\frac{\partial l}{\partial\lambda})_{\hat{\lambda}} = 0, \qquad E(\frac{\partial l}{\partial\gamma})_{\hat{\gamma}} = 0.$$

By simple calculation, we can prove that:

$$\begin{aligned} E(\frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\beta'}) &= -E(\frac{\partial^2 l}{\partial\beta\partial\beta'}), \quad E(\frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\lambda'}) = -E(\frac{\partial^2 l}{\partial\beta\partial\lambda'}), \\ E(\frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\gamma'}) &= -E(\frac{\partial^2 l}{\partial\beta\partial\gamma'}), \quad E(\frac{\partial l}{\partial\lambda}\frac{\partial l}{\partial\lambda'}) = -E(\frac{\partial^2 l}{\partial\lambda\partial\lambda'}), \\ E(\frac{\partial l}{\partial\lambda}\frac{\partial l}{\partial\gamma'}) &= -E(\frac{\partial^2 l}{\partial\lambda\partial\gamma'}), \quad E(\frac{\partial l}{\partial\gamma}\frac{\partial l}{\partial\gamma'}) = -E(\frac{\partial^2 l}{\partial\gamma\partial\gamma'}). \end{aligned}$$

Then, similar to the proof in [13], we have the asymptotic normality of estimators.

Theorem 2. *Based on the regularity conditions and Liapounov form of the multivariate central limit theorem,* $(\hat{\beta}', \hat{\lambda}', \hat{\gamma}')'$ *is asymptotically normally distributed with variance* $I(\beta_0, \lambda_0, \gamma_0)^{-1}$*:*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\lambda}-\lambda_0\\ \hat{\gamma}-\gamma_0\end{pmatrix} \xrightarrow{d} MVN(\vec{0}, I(\beta_0,\lambda_0,\gamma_0)^{-1}).$$

Theorems 1 and 2 indicate the asymptotic consistency and normality of AHPC estimators $(\hat{\beta}', \hat{\lambda}', \hat{\gamma}')'$, respectively. Regularity conditions and the proof of theorems above are included in Supplementary Materials.

6. Example Studies

In this section we set up several numerical examples to test the order-dependence of HPC and permutation invariance of AHPC. We directly decompose and model the correlation matrix R_0 using both methods instead of making estimations by fitting sample data. In this way we isolate the permutation variation, which is the main difference between the two methods, from randomness.

6.1. Order-Dependency of Entries in Φ of HPC

Giving a set of Y_j $j = 1, 2, 3, 4, 5, 6$, assume there is no nature order for Y_js. Then, under order $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$, we will have HPC decomposition of R_0 as $R_0 = T_0 T_0'$ and the corresponding Φ_0 matrix,

$$T_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0\\ 0.980 & 0.198 & 0 & 0 & 0 & 0\\ 0.921 & 0.321 & 0.219 & 0 & 0 & 0\\ 0.696 & 0.387 & 0.218 & 0.562 & 0 & 0\\ 0.169 & -0.028 & -0.223 & -0.399 & 0.872 & 0\\ -0.588 & -0.596 & -0.467 & -0.265 & -0.093 & 0.013 \end{pmatrix}, \Phi_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0\\ 0.2 & 0 & 0 & 0 & 0 & 0\\ 0.4 & 0.6 & 0 & 0 & 0 & 0\\ 0.8 & 1.0 & 1.2 & 0 & 0 & 0\\ 1.4 & 1.6 & 1.8 & 2.0 & 0 & 0\\ 2.2 & 2.4 & 2.6 & 2.8 & 3.0 & 0 \end{pmatrix}.$$

First, we change the order into $Y_1, Y_2, Y_5, Y_4, Y_3, Y_6$ by switching the position between Y_3 and Y_5. That will make R_0 into $R_{3/5}$ by relocating entries' positions in the dark shaded area:

$$R_0 \begin{pmatrix} 1 & 0.980 & 0.921 & 0.696 & 0.169 & -0.588\\ 0.980 & 1 & 0.966 & 0.759 & 0.160 & -0.695\\ 0.921 & 0.966 & 1 & 0.814 & 0.098 & -0.836\\ 0.696 & 0.759 & 0.814 & 1 & -0.166 & -0.892\\ 0.169 & 0.160 & 0.098 & -0.166 & 1 & 0.046\\ -0.588 & -0.695 & -0.836 & -0.892 & 0.046 & 1 \end{pmatrix}, R_{3/5} \begin{pmatrix} 1 & 0.980 & 0.169 & 0.696 & 0.921 & -0.588\\ 0.980 & 1 & 0.160 & 0.759 & 0.966 & -0.695\\ 0.169 & 0.160 & 1 & -0.166 & 0.098 & 0.046\\ 0.696 & 0.759 & -0.166 & 1 & 0.814 & -0.892\\ 0.921 & 0.966 & 0.098 & 0.814 & 1 & -0.836\\ -0.588 & -0.695 & 0.046 & -0.892 & -0.836 & 1 \end{pmatrix}$$

The HPC decomposition of $R_{3/5}$ is as follows:

$$T_{3/5} = \left(\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0\\ 0.980 & 0.198 & 0 & 0 & 0 & 0\\ 0.169 & -0.028 & 0.985 & 0 & 0 & 0\\ 0.696 & 0.387 & -0.277 & 0.535 & 0 & 0\\ 0.921 & 0.321 & -0.049 & 0.063 & 0.204 & 0\\ -0.588 & -0.596 & 0.131 & -0.401 & -0.345 & 0.013 \end{array}\right), \Phi_{3/5} = \left(\begin{array}{cc|cccc} 0.0 & 0.0 & 0 & 0 & 0 & 0\\ 0.2 & 0.0 & 0 & 0 & 0 & 0\\ 1.4 & 1.6 & 0 & 0 & 0 & 0\\ 0.8 & 1.0 & 2.048 & 0 & 0 & 0\\ 0.4 & 0.6 & 1.8 & 1.268 & 0 & 0\\ 2.2 & 2.4 & 1.328 & 2.429 & 3.103 & 0 \end{array}\right).$$

The values of entries on the left-hand side of the vertical line remain the same. Only the location of the entries changes accordingly. For the reasons we stated before, on the right-hand side are all new values, except 0s. This simple simulation study testifies that the HPC method on the correlation matrix is order dependent in terms of entries in Φ. Further numeric example studies are in Supplementary Materials.

6.2. Order Dependency of Covariance Model of HPC

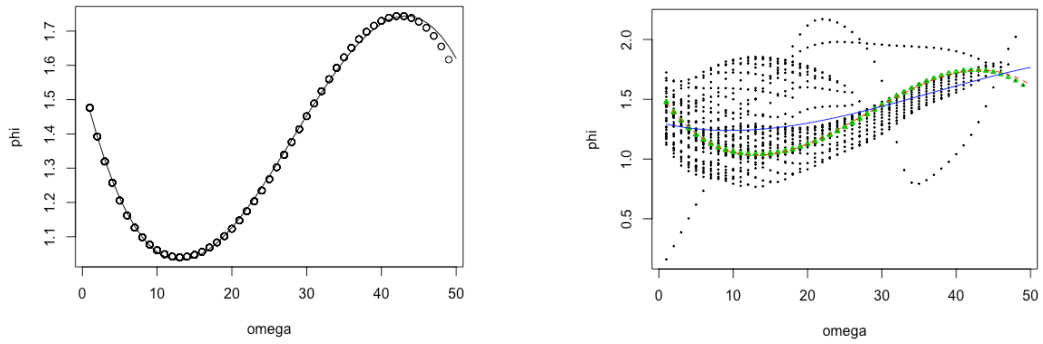
Next, we study how the change of order under HPC would affect model assumption. Setting dimension $p = 50$, and a cubic linear model for correlation component ϕ_{jk}:

$$\phi_{jk} = \gamma_1 + \gamma_2 \omega_{jk} + \gamma_3 \omega_{jk}^2 + \gamma_4 \omega_{jk}^3, \tag{32}$$

where $\omega_{jk} = |j-k|$ $(1 < k < j \leq p)$. Under the starting sequential order $1, 2, \ldots, 50$, $\gamma_{null} = (1.671, -1.004 \times 10^{-03}, -4.982 \times 10^{-04}, 9.967 \times 10^{-06})'$. Based on model (32) and γ_{null} we can have Φ_{null}, T_{null} and R_{null}. We plot ϕ_{jk}s out against their corresponding ω_{jk}s; we can see they form a smooth cubic line in Figure 3a. Then we switch the order 1 and 50; under the new order $50, 2, 3, \ldots, 49, 1$ we can have the new correlation matrix $R_{1/50}$ by re-allocating some relevant elements. By applying the HPC method, we have $T_{1/50}$ and $\Phi_{1/50}$. Then we rearrange the design matrix $\Omega_{1/50}$ according to the new order. Again we

plot $\Phi_{1/50}$ against their corresponding $\Omega_{1/50}$ in Figure 3b. It can be observed that all values are different since the order changed.

We fit a new cubic liner model on the ϕ_{jk}s under the new order. As a result, new $\hat{\gamma}_{1/50}$ is generated, $\hat{\gamma}_{1/50} = (1.614594, 1.503461 \times 10^{-03}, -3.527650 \times 10^{-04}, 5.440070 \times 10^{-06})'$. It is obvious that $\hat{\gamma}_{1/50}$ in Figure 3a is quite different from the original γ_{null}, which is plotted in Figure 3b.



(**a**) (**b**)

Figure 3. Plot of ϕ_{jk}s against their corresponding ω_{jk}s. (**a**) Plot of ϕ_{jk}s against their corresponding ω_{jk}s, with γ_{null} under sequential order $1, \ldots, 50$, and the solid line in the true model. (**b**) Plot of ϕ_{jk} against ω_{jk} under new order $50, 2, \ldots, 49, 1$ in dots. The solid line is the fitted line with $\hat{\gamma}_{1/50}$, while the dotted line stands for the γ_{null} model.

Even worse, when using $\hat{\gamma}_{1/50}$ to create an estimate $\hat{R}_{1/50}$, we find $\hat{R}_{1/50}$ is a poor estimator of $R_{1/50}$ with a large sum of the absolute difference (**SAD**), the Frobenius norm $||R_{1/50} - \hat{R}_{1/50}|| = 60.91713$. That implies that the model assumption based on the original order may not even be appropriate under the new order. Furthermore, to test how the HPC method reacts to different kinds of order changes, we do a series of simulations as described below.

Algorithm 3: HPC Reactions to Different Permutation Changes

1 Switch order j and k. And get real $R_{j/k}$ by reallocating elements in R_{null}.
2 Apply HPC method on $R_{j/k}$ to get true value for $T_{j/k}$ and $\Phi_{j/k}$. And rearrange Ω into $\Omega_{j/k}$ under the changed order.
3 Fit a cubic liner model on $\Phi_{j/k}$ against $\Omega_{j/k}$ to get estimator $\hat{\gamma}_{j/k}$. And use $\hat{\gamma}_{j/k}$ to create estimator $\hat{\Phi}_{j/k}$, $\hat{T}_{j/k}$ and $\hat{R}_{j/k}$.
4 Record the Frobenius norms $||R_{j/k} - \hat{R}_{j/k}||$, $||T_{j/k} - \hat{T}_{j/k}||$ and $||\Phi_{j/k} - \hat{\Phi}_{j/k}||$.

First, we test how the distance between the pair of data we switch would affect the **SAD** between the estimator and true value. Here we do 49 times different switches between order 1 and k ($k = 2, 3, \ldots, 50$). The results can be seen in Figure 4, from which we can observe the **SAD** between estimators and the true value getting bigger with the increase of distance between switched positions. That indicates that the further position being switched, the more significantly biased the estimator becomes.

Second, we test how the first position we switch would affect the estimators. This time we also use the procedure above, while setting $j = 1, 2, 3, \ldots, 49$ and $k = j + 1$. Then we have Figure 5, from which we can see the bias decrease as further change happens from the first order. Furthermore, comparing with Figure 4, we can see the effect of distance between the pair of switched positions is much stronger than that of the minimum number being changed.

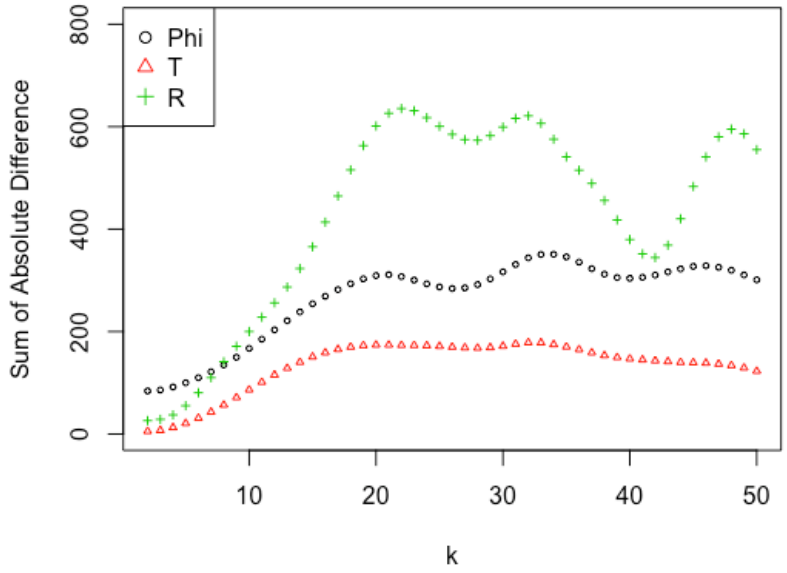
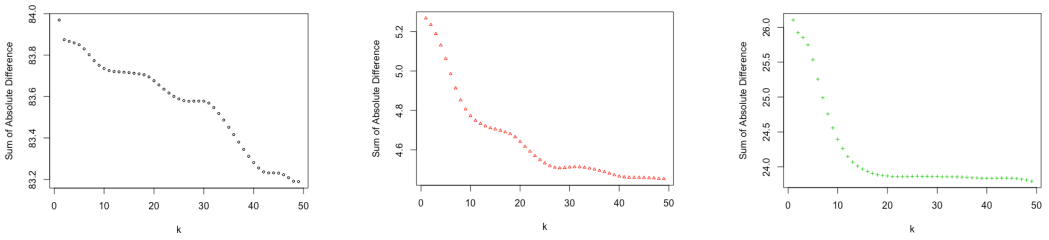


Figure 4. Plot of the difference between estimators and true value when switching order 1 and k ($k = 2, 3, \ldots, 50$).



(**a**) $sum(|\Phi_{est} - \Phi_{true}|)$. (**b**) $sum(|T_{est} - T_{true}|)$. (**c**) $sum(|R_{est} - R_{true}|)$.

Figure 5. Plot of differences before and after switching order j and $j+1$ ($j = 1, 2, 3, \ldots, 49$).

6.3. Permutation Invariance of Entries in Φ in AHPC

Now we apply the same simulation study on AHPC. Giving the same starting correlation matrix R_0 as above, after Cholesky decomposition on R_0, the same T_0 as the one before is generated. Based on Equation (14) we have $\tilde{\Phi}_0$ for the AHPC. Then we swap positions 3 and 5, which gives $R_{3/5}$ and $T_{3/5}$ as those in HPC. According to the new proposed definition of the translation between Φ and T in AHPC, we have $\tilde{\Phi}_{3/5}$ for AHPC:

$$\tilde{\Phi}_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.259 & 0 & 0 & 0 & 0 \\ 0.8 & 0.707 & 0.619 & 0 & 0 & 0 \\ 1.4 & 1.409 & 1.472 & 1.737 & 0 & 0 \\ 2.2 & 2.339 & 2.561 & 2.674 & 1.524 & 0 \end{pmatrix}, \tilde{\Phi}_{3/5} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0 \\ 1.4 & 1.409 & 0 & 0 & 0 & 0 \\ 0.8 & 0.707 & 1.737 & 0 & 0 & 0 \\ 0.4 & 0.259 & 1.472 & 0.619 & 0 & 0 \\ 2.2 & 2.339 & 1.524 & 2.674 & 2.561 & 0 \end{pmatrix}.$$

all those gray shaded parts have positions re-allocated, but the value remains the same as the original $\tilde{\Phi}_0$ for AHPC. We can see that, after the order changes, all values in matrix Φ under AHPC remain the same; only their position changes to fit the new order. In that sense, we say AHPC is order invariant in terms of matrix Φ.

6.4. Permutation Invariance of Covariance Model in AHPC

Then to find out how the AHPC modeling reacts to the order change, we implement the same γ_{null} and cubic linear model assumption (32) under sequential order $1, 2, \ldots, 50$. Applying the AHPC method and under the model assumption, we have Φ^*_{null}, T^*_{null} and R^*_{null}. After switching the order between 1 and 50, applying AHPC we have $R^*_{1/50}$, $T^*_{1/50}$ and $\Phi^*_{1/50}$. This time we can see all values in $\Phi^*_{1/50}$ are as same as those in Φ^*_{null}. Plotting ϕ_{jk}s against their corresponding ω_{jk}s, we have Figure 6.

Furthermore, we fit a cubic linear model to ϕ_{jk} that, under the new order, against their corresponding ω_{jk}s, involves entries from $\Omega_{1/50}$. Then we get an estimator $\hat{\gamma}_{1/50} = \gamma_{null}$. Moreover, the $\hat{R}^*_{1/50}$ also has zero bias from $R^*_{1/50}$. In conclusion, the AHPC is order invariant in terms of both value in Φ^* and covariance modeling.

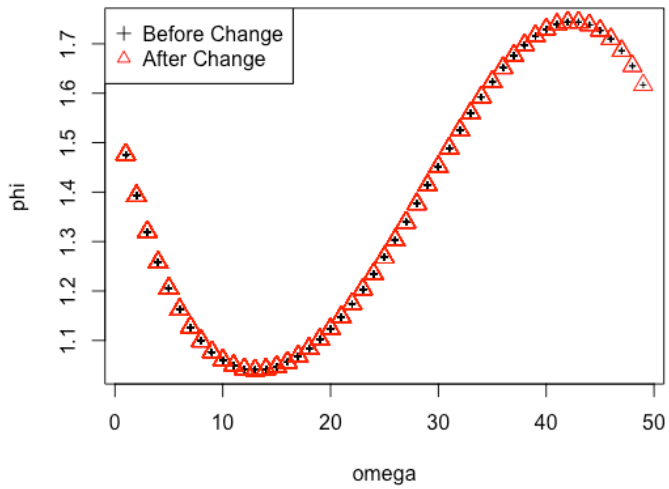
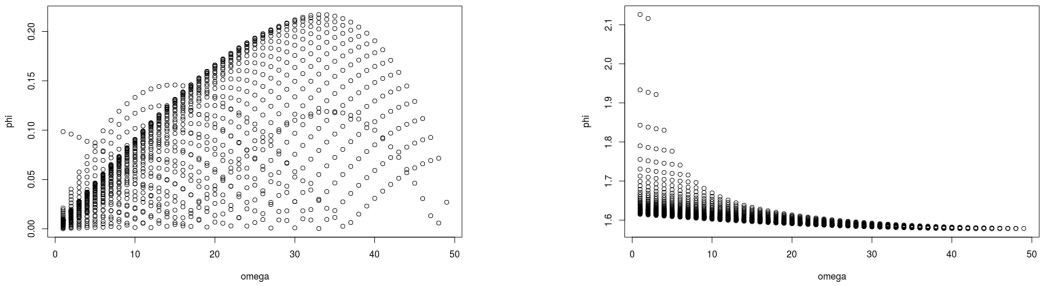


Figure 6. Plots of ϕ^*_{jk}s in the AHPC method against ω_{jk}s under the original sequential order and switched order.

The model assumptions under HPC and AHPC are not compatible. Continue the simulation above; under the cubic linear model assumption, when correlation matrix R_{HPC} is generated from HPC, we decompose R_{HPC} by AHPC and end up with a plot of ϕ^*_{jk} against ω_{jk} in Figure 7a. The other way around, when correlation matrix R_{AHPC} is created by AHPC we decompose R_{AHPC} by HPC and plot Figure 7b. As we can see, the model assumptions in HPC and AHPC are not compatible with each other. Based on the asymptotic research, the consistency only holds when the model assumption is correct. Thus there is no point to doing the cross method estimation comparison under one identical model assumption.



(**a**) (**b**)

Figure 7. Cross method results. (**a**) ϕ_{jk}s against ω_{jk}s from R_{HPC} de-composited by the AHPC method. (**b**) ϕ_{jk}s against ω_{jk}s from R_{MHPC} de-composited by the HPC method.

7. Real Data Analysis

In this section, we analyze a set of weather data by HPC and AHPC. We focus on comparing the estimators of the correlation matrix. Furthermore, there is no natural order in this weather data set. In this section, **jmcm** package in **R** [5] is used for estimations of the HPC method. Function ***lm*** is applied in **R** to do the linear regression for AHPC. These weather data are collected from the UK government public web page, **Met Office**. We select $p = 5$ different weather stations in the UK. They are allocated in Ballypartick Forest,

Cambridge, Lewick, Leuchars and Sheffield. We cut off the data before 1962 and make this data set balance, leaving us with $n = 660$ sample size. There is no missing value in this data set. The average minimum temperature in each month is recorded in Celsius degrees.

For comparison between HPC and AHPC, our interest is on the estimator of correlation matrices. Thus, sample variance $\tilde{\sigma}_i$ is used as the estimator of the variance for station i. We set the initial order alphabetically: Ballypartick Forest, Cambridge, Lewick, Leuchars and Sheffield. Then we shift this initial order by relocating the Sheffield station to the beginning. Meanwhile, based on the non-parametric analysis of Φ under both methods with respect to distances d_{ij} between climate stations i and j, solid lines are plotted in Figure 8. Combined with the consideration of **AIC** and **BIC**, we assume three different polynomial models for ϕ in HPC and AHPC under two orders, respectively.

For ϕ_{ij}s of the HPC method, we assume a linear model under the initial order and a quadratic one under the shifted order. On the other hand, to keep the model monotonic decreasing, we model angle parameters $(1/\phi_{ij})$s in AHPC instead of ϕ_{ij}s. We assume a linear model under both orders.

Under both orders, by applying HPC and AHPC, we can get four regression results. The fitted models are plotted with dot lines in Figure 8. Cross comparison of these results with the sample correlation matrices under these two orders, by their relative errors defined as Relative Error $= ||R_S - \hat{R}||/||R_S||$, where $||.||$ is Euclidean norm, and R_S and $\hat{R}$ are the sample and estimated correlation matrix, respectively. We should notice that the norm of sample correlations remains the same, $||R_{initial}|| = ||R_{shifted}|| = 11.055$, and the norm of the difference between sample correlation matrices under both is order $||R_{initial} - R_{shifted}|| = 13.69$.

Observing the results in Tables 2 and 3, it is obvious that the estimator of HPC depends on the order of the data, while AHPC presents a consistent estimating result. Even redoing the modeling process from the model selecting stage, we observe from the relative errors in Table 2 that the relative errors between sample and estimated correlation generated on $\gamma_{initial}$&$\gamma_{shifted}$ are different in HPC. As one conclusion, under different permutations, the estimating results of HPC would vary. Consequentially, it is improper to use the HPC method to model the covariance matrices of data without a natural order for its failure to offer a permutation consistent estimator.

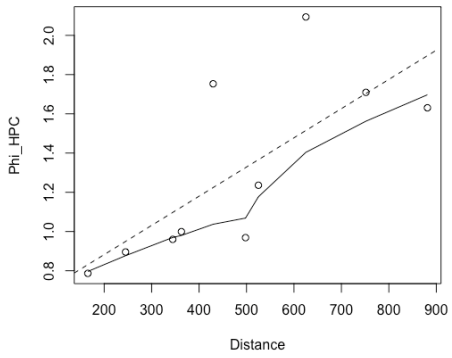
Table 2. Resulting HPC estimates and standard error of correlation components γ, under initial and shifted order, respectively. Relative errors between sample and estimated correlation under different orders are cross compared.

Data Order	**Initial Order**		**Shifted Order**	
	Estimator	Standard Error	Estimator	Standard Error
γ_0	5.847	2.598×10^{-1}	-1.544×10^{-1}	5.672×10^{-1}
γ_1	1.488×10^{-3}	4.93×10^{-4}	5.186×10^{-3}	2.375×10^{-3}
γ_2	0	0	-3.480×10^{-6}	2.222×10^{-6}
	$\hat{R}(\hat{\gamma}_{initial})$	$\hat{R}(\hat{\gamma}_{shifted})$	$\hat{R}(\hat{\gamma}_{initial})$	$\hat{R}(\hat{\gamma}_{shifted})$
Relative Error	0.173	0.216	1.575	0.124

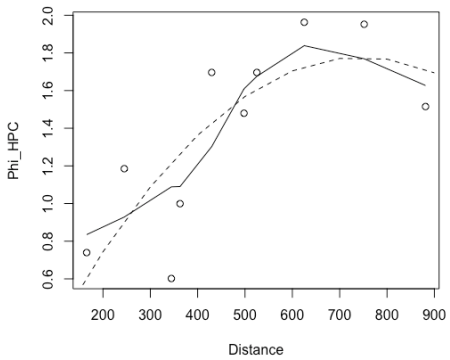
On the other hand, for AHPC, Table 3 shows the consistency of these estimators under different orders. Moreover, there is an interpretation advantage in AHPC comparing to HPC. For example, in this real data analysis, the interpretation of the relationship between correlation ρ_{jk} and the distance between stations j and k are not obvious in the HPC model. Meanwhile, the AHPC model for $1/\phi_{ij}$ is monotonic decreasing, suggesting ϕ_{jk} increases with distance ω_{jk}. Moreover, $\rho_{jk} = cos(\phi_{jk})$ $\phi_{jk} \in (0, \pi]$ is a monotonic decreasing function. Thus ρ_{jk} decreases with respect to the distance between stations.

Table 3. Resulting AHPC estimates and standard error of correlation components γ, under initial and shifted order, respectively. Relative errors between sample and estimated correlation under different orders are cross compared.

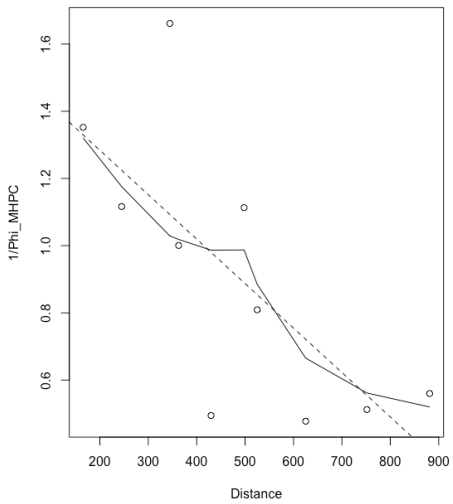
Data Order	**Initial Order**		**Shifted Order**	
	Estimator	Standard Error	Estimator	Standard Error
γ_0	1.548	2.365×10^{-1}	1.548	2.365×10^{-1}
γ_1	-1.323×10^{-3}	4.488×10^{-4}	-1.323×10^{-3}	4.488×10^{-4}
	$\hat{R}(\hat{\gamma}_{initial})$	$\hat{R}(\hat{\gamma}_{shifted})$	$\hat{R}(\hat{\gamma}_{initial})$	$\hat{R}(\hat{\gamma}_{shifted})$
Relative Error	0.347	0.347	0.347	0.347



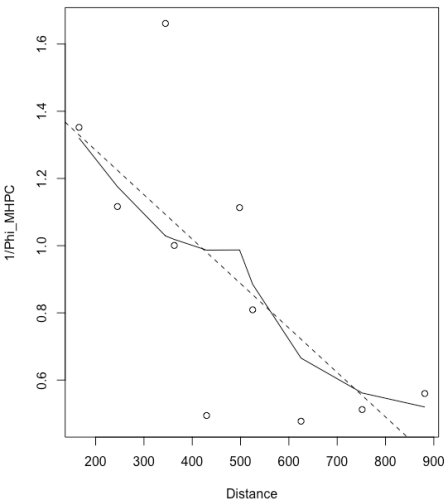
(**a**) HPC-Origin



(**b**) HPC-Shifted



(**c**) AHPC-Shifted



(**d**) AHPC-Origin

Figure 8. Solid and dotted lines are non-parametric smooth and fitted model of covariance components Φ (circles in each plots) under HPC and AHPC methods separately with respect to distances between five stations, under initial and shifted order.

8. Conclusions and Discussion

In this paper, we addressed the permutation variation of HPC through its geometrical interpretations. Then AHPC for covariance modeling was proposed. AHPC was proven to improve the order-dependence issue of HPC. Furthermore, the direct relation between Φ and R in AHPC provides an advantage in making model assumptions, parameter estimations and statistical interpretations. However, due to the limitation of the relation between angles, the model assumption for angles Φ in AHPC must satisfy an extra constraint.

Both HPC and AHPC can only guarantee semi-positive definiteness. The reason behind this drawback of both methods is the same. From the geometrical interpretations above, we can see that the definition of angles in both methods can only ensure the symmetry and diagonal elements being 1s in the correlation matrix R. By changing the inequality constraint on angles in AHPC to strict less, we can simultaneously make sure the correlation matrix is positive definite.

These four covariance modeling methods we mentioned in this paper have their advantages. For the most accurate estimator, an appropriate model assumption is essential. Thus we may use a certain decomposition method based on the data, since different methods would generate their pattern against coefficients. Under an appropriate model assumption, estimators in all four methods are consistent. AHPC, MCD and ACD can do the model selection visually, while in some cases, model selection in HPC can only be based on statistical criteria.

There are some potential researches available on this AHPC method. As we see in simulation examples, sometimes the pattern is too complicated to fit with the linear model; the non-parametric and semi-parametric model could be applied in the AHPC method, similar to the studies of [14,15].

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/math10040562/s1.

Author Contributions: Conceptualization, J.P. and Q.L.; methodology, J.P. and Q.L.; software, Q.L.; validation, J.P.; writing—original draft preparation, Q.L.; writing—review and editing, Q.L.; visualization, Q.L.; supervision, J.P.; project administration, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant number 11731011 and 11871357.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pourahmadi, M. *High-Dimensional Covariance Estimation*; Wiley: Hoboken, NJ, USA, 2013.
2. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Volume 793,
3. Zhang, C.H.; Huang, J. The sparsity and bias of the lasso selection in high-dimensional linear regression. *Ann. Stat.* **2008**, *36*, 1567–1594. [CrossRef]
4. Pourahmadi, M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **1999**, *86*, 677–690. [CrossRef]
5. Pan, J.; Pan, Y. jmcm: An R package for joint mean-covariance modeling of longitudinal data. *J. Stat. Softw.* **2017**, *82*, 1–29. [CrossRef]
6. Rebonato, R.; Jäckel, P. The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. 2011. Available online: https://ssrn.com/abstract=1969689 (accessed on 8 May 2012).
7. Rapisarda, F.; Brigo, D.; Mercurio, F. Parameterizing correlations: A geometric interpretation. *IMA J. Manag. Math.* **2007**, *18*, 55–73. [CrossRef]
8. Pourahmadi, M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **2000**, *87*, 425–435. [CrossRef]
9. Murty, K.G. *Computational and Algorithmic Linear Algebra and n-Dimensional Geometry*; World Scientific Publishing Company: Singapore, 2014.
10. Zhang, W.; Leng, C.; Tang, C.Y. A joint modelling approach for longitudinal studies. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2015**, *77*, 219–238. [CrossRef]
11. Bates, D.M.; Watts, D.G. *Nonlinear Regression Analysis and Its Applications*; Wiley: New York, NY, USA, 1988; Volume 2,
12. Sallan, J.M.; Lordan, O.; Fernandez, V. *Modeling and Solving Linear Programming with R*; OmniaScience: Barcelona, Spain, 2015.
13. Chiu, T.Y.; Leonard, T.; Tsui, K.W. The matrix-logarithmic covariance model. *J. Am. Stat. Assoc.* **1996**, *91*, 198–210. [CrossRef]

14. Wang, N.; Carroll, R.J.; Lin, X. Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Am. Stat. Assoc.* **2005**, *100*, 147–157. [CrossRef]
15. Fan, J.; Huang, T.; Li, R. Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Am. Stat. Assoc.* **2007**, *102*, 632–641. [CrossRef] [PubMed]