Supplementary Materials: Permutation Variation and Alternative Hyper-Sphere Decomposition

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

1. Asymptotic Properties in Details

In this section, consistency and asymptotic normality of **MHPC** estimators under standard situation $p < n$ are proven. Similar to the proof in Zhang and et al. (2004), under some regularity conditions:

1. Model assumptions for mean and variance-covariance components are correct.
2. The dimensions q_β, q_λ and q_γ for parameters β, λ and γ are fixed, meanwhile sample size $n \to \infty$.
3. The parameter space for β, λ and γ are compact subspaces of R^{q_β}, R^{q_λ} and R^{q_γ} respectively. And the true values β_0, λ_0 and γ_0 are in the interiors of corresponding subspaces.
4. As sample size $n \to \infty$, $n^{-1}I$ almost surely converges to a positive definite matrix $\mathcal{I}$, where I is the negative expected Hessian matrix .

Here the compactness in **Condition 3** means the solution subspaces for our estimators are closed, and also ensures the distance between any estimator and true value is bounded, which means there exists a ϵ so that

$$||\hat{\beta} - \beta_0|| < \epsilon, \qquad ||\hat{\lambda} - \lambda_0|| < \epsilon, \qquad ||\hat{\gamma} - \gamma_0|| < \epsilon.$$

Then we can have the asymptotic consistency of estimators $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\gamma}$.

Theorem 1. *Under the regularity conditions **1,2,3**, 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. And 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 as the sample size n goes to ∞, it is straightforward to have:

$$\frac{1}{n} I(\hat{\beta}, \hat{\lambda}, \hat{\gamma}) - \frac{1}{n} E(\mathcal{I}) \xrightarrow{a.s} 0, \tag{1}$$

where $\mathcal{I}$ is a positive definite matrix, which we defined in **Condition ??**. And it is easy to prove $\mathcal{I} = I(\beta_0, \lambda_0, \gamma_0)$. Then similar to the proof of Theorem 2 in Chiu et al. (1996) we can

have the asymptotic normality of our estimators.

Theorem 2. *Based on the **Condition 1,2,3,4** and Liapounov form of the multivariate central limit theorem,* $(\hat{\beta}', \hat{\lambda}', \hat{\gamma}')'$ *is asymptotically normally distributed with variance* $I(\beta_0, \lambda_0, \gamma_0)^{-1}$*:*

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

Theorem 1 and 2 indicate the asymptotic consistency and normality of **MHPC** estimators $(\hat{\beta}', \hat{\lambda}', \hat{\gamma}')'$ separately.

The prove is similar as in Chui et al. (1996). Giving a set of estimators $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\gamma}$, and setting

$$r_i = y_i - X\beta_0 + X(\beta_0 - \hat{\beta}), \tag{2}$$

where β_0 is the true value for coefficient β in mean model. Then we have

$$E(r_i) = X(\beta_0 - \hat{\beta}), \tag{3}$$

and

$$E(r_i r_i') = \Sigma_0 + X(\beta_0 - \hat{\beta})(\beta_0 - \hat{\beta})'X', \tag{4}$$

where Σ_0 is the true covariance matrix build with λ_0 and γ_0. Giving a sample y_i ($i = 1, 2, \ldots, n$), ignoring the constant part we have the log-likelihood function $l_i = \log f(y_i, \beta, \lambda, \gamma)$ as follow:

$$-2l_i = \log|\hat{\Sigma}| + r_i'\hat{\Sigma}^{-1}r_i \tag{5}$$

When $\beta = \hat{\beta}$, $\lambda = \hat{\lambda}$ and $\gamma = \hat{\gamma}$ we have the mean and variance of the log-likelihood for sample y_i as:

$$\begin{array}{ll} -2E(l_i) & = \log|\hat{\Sigma}| + tr(\hat{\Sigma}^{-1}\Sigma_0) + (\beta_0 - \hat{\beta})'X'\hat{\Sigma}^{-1}X(\beta_0 - \hat{\beta}), \\ 2var(l_i) & = tr(\hat{\Sigma}^{-1}\Sigma_0)^2 + 2\big[(\beta_0 - \hat{\beta})'X'\hat{\Sigma}^{-1}\Sigma_0\hat{\Sigma}^{-1}X(\beta_0 - \hat{\beta})\big]. \end{array} \tag{6}$$

Based on the compactness of solution subspaces and boundedness of the covariates it can be shown that there exist a constant κ so that for all $1 \le i \le n$, $var(l_i) \le \kappa$. Therefore by the Kolmogorov's strong law of large number average of the log-likelihood l_i is almost surely convergence to the average of its mean:

$$\frac{1}{n}\sum_{i=1}^{n} l_i - \frac{1}{n}\sum_{i=1}^{n} E(l_i) \xrightarrow{a.s.} 0, \qquad \text{as } n \to \infty. \tag{7}$$

The κ in above is independent of the estimators, so the average of the expected log-likelihood $\frac{1}{n}\sum_{i=1}^{n} E(l_i)$ is equi-continuous in solution subspaces. Then similar to the proof of Theorem 1 in Chiu et al. (1996), we can also prove the consistency of our estimators here.

2. Derivative of Score and Expected Fisher Information Matrices

$$I_{\lambda\lambda'} = -E(\frac{\partial^2 l}{\partial\lambda\partial\lambda'}) = \frac{n}{4}\sum_{l=1}^{p}\sum_{k_1=1}^{l}\sum_{k_2=1}^{l} a_{lk_1}a_{lk_2}\rho_{k_1k_2}(h_{k_1}h_{k_1}' + h_{k_2}h_{k_1}'). \tag{8}$$

Based on some properties of derivative of matrix:

$$\frac{\partial}{\partial X}tr(AXB) = A'B', \tag{9}$$

and

$$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1}\frac{\partial Y}{\partial x}Y^{-1}, \tag{10}$$

where $Y = g(x)$, we can have:

$$\begin{aligned}\frac{\partial}{\partial \gamma_m} r_i' D^{-1} R^{-1} D^{-1} r_i &= \frac{\partial}{\partial R^{-1}} tr(r_i' D^{-1} R^{-1} D^{-1} r_i)\frac{\partial R^{-1}}{\partial \gamma_m} \\ &= tr(D^{-1} r_i r_i' D^{-1}(-R^{-1})\Delta(\frac{\partial \rho_{jk}}{\partial \gamma_m})R^{-1}).\end{aligned} \tag{11}$$

Based on:

$$\frac{\partial log|X|}{\partial x} = tr(X^{-1}\frac{\partial X}{\partial x}), \tag{12}$$

we have:

$$\frac{\partial log|DRD|}{\partial \gamma_m} = tr(R^{-1}\frac{\partial R}{\partial \gamma_m}). \tag{13}$$

Similar as in Zhang and Leng (2015),based on the notation of ξ_{i*l} we will have $T_i\xi_{il} = D_i^{-1}r_{il}$. Then we can show the jth element will have equation as:

$$\xi_j = \sum_{k=1}^{j}\frac{a_{jk}}{\sigma_k}r_j, \tag{14}$$

Then taking the first derivative of ξ_j with respect to λ is:

$$\partial\xi_j/\partial\lambda = \sum_{k=1}^{j} a_{jk} r_k \partial\sigma_k^{-1}/\partial\lambda, \tag{15}$$

and

$$\partial\sigma_k^{-1}/\partial\lambda = -\frac{1}{2}h_k/\sigma_k. \tag{16}$$

Then taking the first derivative of our log-likelihood with respect to λ will give us the score function for λ:

$$\begin{aligned}U_\lambda = \partial l/\partial\lambda &= -\frac{1}{2}n\textstyle\sum_{l=1}^{p}\frac{\partial log\sigma_l^2}{\partial\lambda} - \frac{1}{2}\sum_{l=1}^{p}\sum_{j=1}^{n}\frac{\partial\xi_j^2}{\partial\lambda} \\ &= -\frac{1}{2}n\textstyle\sum_{l=1}^{p}h_l + \frac{1}{2}\sum_{l=1}^{p}\sum_{j=1}^{n}\xi_j\sum_{k=1}^{j}a_{jk}r_k h_k/\sigma_k.\end{aligned} \tag{17}$$

Taking the second derivative of log-likelihood with respect to λ_i

$$\begin{aligned}-2\frac{\partial^2 l}{\partial\lambda_i\partial\lambda_i'} &= -\textstyle\sum_{j=1}^{n}\sum_{l=1}^{p}\left[(\sum_{k=1}^{l}a_{lk}\frac{r_{kj}}{\sigma_k}h_k\frac{\partial\xi_j}{\partial\lambda'}) + (\xi_j\sum_{k=1}^{l}a_{lk}r_k h_k\frac{\partial\sigma_k^{-1}}{\partial\lambda'})\right] \\ &= -\textstyle\sum_{j=1}^{n_i}\sum_{l=1}^{p}\Big[(-\frac{1}{2}\sum_{k_1=1}^{l}a_{lk_1}\frac{r_{k_1}}{\sigma_{k_1}}h_{k_1}\sum_{k_2=1}^{l}a_{lk_2}\frac{r_{k_2}}{\sigma_{k_2}}h_{k_2}') \\ &\quad +(-\frac{1}{2}\xi_{lj}\sum_{k=1}^{l}a_{lk}\frac{r_k}{\sigma_k}h_k h_k')\Big].\end{aligned} \tag{18}$$

Giving that $\frac{\partial\xi_j}{\partial\lambda}$ is known, taking the second derivative of it with respect to λ'. And with equations (16), (15), as the result the expectation of (18) is (8).

Since $E(r_i) = \vec{0}$ the expectation of $I_{\gamma_m\beta}$:

$$\begin{aligned} E\left[I_{\gamma_m\beta}\right] = E\left[I_{\beta\gamma_m}\right] &= -1/2\sum_{i=1}^{n} E\left[\left(\frac{\partial r_i'}{\partial\beta}D^{-1}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}r_i\right)\right. \\ &\left.\left(r_i'D^{-1}R^{-1}\Delta(\frac{\partial\rho_{jk}}{\partial\gamma_m})R^{-1}D^{-1}\frac{\partial r_i}{\partial\beta}\right)\right] \\ &= \vec{0} \end{aligned} \quad . \tag{19}$$

Based on same reason, the off diagonal blocks in expected Fisher-information matrix between β and variance component λ also is **0**.

$$\begin{aligned} 2I_{\beta\lambda'} = 2I'_{\lambda\beta'} &= -2E(\frac{\partial^2 l}{\partial\beta\partial\lambda'}) \\ &= X'\frac{\partial}{\partial\lambda}\Sigma_i^{-1}E(\epsilon_i)1_{n_i} = \mathbf{0}, \end{aligned} \tag{20}$$

3. More Example Studies

3.1. Order-dependency of HPC

We assume a set of Y_j $j = 1, 2, 3, 4, 5, 6$. There is no nature order for Y_js. And correlation matrix for Y_js is:

$$R_0 = \begin{pmatrix} 1 & 0.980 & 0.921 & 0.696 & 0.169 & -0.588 \\ 0.980 & 1 & 0.966 & 0.759 & 0.160 & -0.695 \\ 0.921 & 0.966 & 1 & 0.814 & 0.098 & -0.836 \\ 0.696 & 0.759 & 0.814 & 1 & -0.166 & -0.892 \\ 0.169 & 0.160 & 0.098 & -0.166 & 1 & 0.046 \\ -0.588 & -0.695 & -0.836 & -0.892 & 0.046 & 1 \end{pmatrix} \tag{21}$$

Then under order $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$, we will have HPC decomposition of R_0 as follow:

$$R_0 = T_0T_0'$$

Where T_0 is:

$$T_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.980 & 0.198 & 0 & 0 & 0 & 0 \\ 0.921 & 0.321 & 0.219 & 0 & 0 & 0 \\ 0.696 & 0.387 & 0.218 & 0.562 & 0 & 0 \\ 0.169 & -0.028 & -0.223 & -0.399 & 0.872 & 0 \\ -0.588 & -0.596 & -0.467 & -0.265 & -0.093 & 0.013 \end{pmatrix} \tag{22}$$

And corresponding Φ_0 matrix is:

$$\Phi_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 0.8 & 1.0 & 1.2 & 0 & 0 & 0 \\ 1.4 & 1.6 & 1.8 & 2.0 & 0 & 0 \\ 2.2 & 2.4 & 2.6 & 2.8 & 3.0 & 0 \end{pmatrix}$$

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

$$R_{3/5} = \begin{pmatrix} 1 & 0.980 & 0.169 & 0.696 & 0.921 & -0.588 \\ 0.980 & 1 & 0.160 & 0.759 & 0.966 & -0.695 \\ 0.169 & 0.160 & 1 & -0.166 & 0.098 & 0.046 \\ 0.696 & 0.759 & -0.166 & 1 & 0.814 & -0.892 \\ 0.921 & 0.966 & 0.098 & 0.814 & 1 & -0.836 \\ -0.588 & -0.695 & 0.046 & -0.892 & -0.836 & 1 \end{pmatrix} \tag{23}$$

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

$$T_{3/5} = \left(\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.980 & 0.198 & 0 & 0 & 0 & 0 \\ 0.169 & -0.028 & 0.985 & 0 & 0 & 0 \\ 0.696 & 0.387 & -0.277 & 0.535 & 0 & 0 \\ 0.921 & 0.321 & -0.049 & 0.063 & 0.204 & 0 \\ -0.588 & -0.596 & 0.131 & -0.401 & -0.345 & 0.013 \end{array}\right) \tag{24}$$

Value of entries on the left-hand side of the vertical line remains the same, only the location of entries have changed accordingly. For the reasons we stated before, on the right-hand side are all new values, expect 0s. Then we will have $\Phi_{3/5}$ as:

$$\Phi_{3/5} = \left(\begin{array}{cc|cccc} 0.0 & 0.0 & 0 & 0 & 0 & 0 \\ 0.2 & 0.0 & 0 & 0 & 0 & 0 \\ 1.4 & 1.6 & 0 & 0 & 0 & 0 \\ 0.8 & 1.0 & 2.048 & 0 & 0 & 0 \\ 0.4 & 0.6 & 1.8 & 1.268 & 0 & 0 \\ 2.2 & 2.4 & 1.328 & 2.429 & 3.103 & 0 \end{array}\right)$$

This is just like what we described before. All angles on the left side of the vertical line stay the same, and only the location has changed. Meanwhile, we will have whole new angles on the right-hand side, for the reason we proved before.
We further change the order into $Y_1, Y_2, Y_5, Y_6, Y_3, Y_4$, by switching the position between Y_4 and Y_6. Under the new order, relocate entries in gray shade and get the correlation matrix $R_{3/5||4/6}$ as:

$$R_{3/5||4/6} = \begin{pmatrix} 1 & 0.980 & 0.169 & -0.588 & 0.921 & 0.696 \\ 0.980 & 1 & 0.160 & -0.695 & 0.966 & 0.759 \\ 0.169 & 0.160 & 1 & 0.046 & 0.098 & -0.166 \\ -0.588 & -0.695 & 0.046 & 1 & -0.836 & -0.892 \\ 0.921 & 0.966 & 0.098 & -0.836 & 1 & 0.814 \\ 0.696 & 0.759 & -0.166 & -0.892 & 0.814 & 1 \end{pmatrix} \tag{25}$$

And $T_{3/5||4/6}$ is:

$$T_{3/5||4/6} = \left(\begin{array}{cc|c||ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.980 & 0.198 & 0 & 0 & 0 & 0 \\ 0.169 & -0.028 & 0.985 & 0 & 0 & 0 \\ -0.588 & -0.596 & 0.131 & 0.530 & 0 & 0 \\ 0.921 & 0.321 & -0.049 & -0.181 & 0.113 & 0 \\ 0.696 & 0.387 & -0.277 & -0.405 & -0.349 & 0.024 \end{array}\right) \tag{26}$$

On the left-hand side of the double vertical line is the remain-same-value part of $T_{3/5||4/6}$ comparing with $T_{3/5}$. And the corresponding $\Phi_{3/5||4/6}$ is:

$$\Phi_{3/5||4/6} = \begin{pmatrix} 0.0 & 0.0 & 0 & 0 & 0 & 0 \\ 0.2 & 0.0 & 0 & 0 & 0 & 0 \\ 1.4 & 1.6 & 0 & 0 & 0 & 0 \\ 2.2 & 2.4 & 1.328 & 0 & 0 & 0 \\ 0.4 & 0.6 & 1.8 & 2.584 & 0 & 0 \\ 0.8 & 1.0 & 2.048 & 2.429 & 3.072 & 0 \end{pmatrix}.$$

The gray-shaded part remains the same value as $\Phi_{3/5}$ after the second change. We can see that the remain-same-value pattern remains after we further change the order from $Y_1, Y_2, Y_5, Y_4, Y_3, Y_6$ to $Y_1, Y_2, Y_5, Y_6, Y_3, Y_4$.

These two simple simulation studies testify that the **HPC** method on correlation matrix is order dependent in terms of value of entries in Φ matrix.

3.2. Order invariance of MHPC

Now we apply the same simulation study on **MHPC** and see how the new method react. Giving the same starting correlation matrix R_0 as (21), after Cholesky Decomposition on R_0, the same T_0 as the one before is generated in (22). Based on equation (**??**) we have $\tilde{\Phi}_0$ for the **MHPC** method as:

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

where entries, ϕ_{jk}, is the angle between $\vec{T}_{(j)}$ and $\vec{T}_{(k)}$. For example, as shown above, $\phi_{53} = 1.472$ is the angle between $\vec{T}_{(5)}$ and $\vec{T}_{(3)}$. Then we swap position 3 and 5, which gives $R_{3/5}$ and $T_{3/5}$ as (23) and (24) in **HPC**. According to the new proposed definition of the translation between Φ and T in **MHPC**, we have $\tilde{\Phi}_{3/5}$ for **MHPC** as follow:

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

all those gray shaded parts have their positions re-allocated, but value remains the same as the original $\tilde{\Phi}_0$ for **MHPC**. Further, switch order 4 and 6 on the previous order, we have $R_{3/5||4/6}$ and $T_{3/5||4/6}$ in (25) and (26). And using the **MHPC** method we get $\tilde{\Phi}_{3/5||4/6}$ as below:

$$\tilde{\Phi}_{3/5||4/6} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 0 \\ 1.4 & 1.409 & 0 & 0 & 0 & 0 \\ 2.2 & 2.339 & 1.524 & 0 & 0 & 0 \\ 0.4 & 0.259 & 1.472 & 2.561 & 0 & 0 \\ 0.8 & 0.707 & 1.737 & 2.674 & 0.619 & 0 \end{pmatrix},$$

those dark shaded elements are the new corresponding position-switched ones comparing to $\tilde{\Phi}_{3/5}$. We can see after many order changes all value in matrix Φ under **MHPC** method remain the same, only their position change to fit the new order. In that sense, we say

MHPC is order invariant in terms of matrix Φ.

References

1. Chiu, T.Y.; Leonard, T.; Tsui, K.W. The matrix-logarithmic covariance model. *Journal of the American Statistical Association* **1996**, *91*, 198–210.
2. Zhang, W.; Leng, C.; Tang, C.Y. A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **2015**, *77*, 219–238.