

Additional Proofs for “A Factor-Adjusted Multiple Testing Procedure with Application to Mutual Fund Selection”

Wei Lan and Lilun Du

Southwestern University of Finance and Economics, and Hong Kong University of Science and Technology

Proof of Lemma 5

Proof. We prove the lemma in the following two steps. In the first step, we demonstrate that $\max_{i \in \mathcal{I}_0} |\bar{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| = O_p\{\sqrt{\log N/T}\}$, followed by extending the results to $\tilde{\sigma}_{\boldsymbol{\eta}, ii}$ in the second step.

We firstly show that $\max_{i \in \mathcal{I}_0} |\bar{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| = O_p\{\sqrt{\log N/T}\}$. Note that $\bar{\sigma}_{\boldsymbol{\eta}, ii} = T^{-1} \bar{\boldsymbol{\eta}}_i^\top \bar{\boldsymbol{\eta}}_i$ with $\bar{\boldsymbol{\eta}}^\top = \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{E}^\top$, where $\bar{\mathbf{Z}}$ and $\mathcal{Q}(\bar{\mathbf{Z}})$ are defined in Section 3.1. As a result, we have $\bar{\sigma}_{\boldsymbol{\eta}, ii} = T^{-1} \boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\bar{\mathbf{Z}}) \boldsymbol{\varepsilon}_i$, which leads to

$$\begin{aligned} |\bar{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| &= |T^{-1} \boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\bar{\mathbf{Z}}) \boldsymbol{\varepsilon}_i - T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i + T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii}| \\ &\leq |T^{-1} \boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\bar{\mathbf{Z}}) \boldsymbol{\varepsilon}_i - T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i| + |T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii}|. \end{aligned} \quad (\text{A.1})$$

We then consider the above two parts separately. We firstly consider the second term. Note that $\boldsymbol{\eta}_i$ follows a multivariate normal distribution with mean zero and covariance matrix $\sigma_{\boldsymbol{\eta}, ii} \mathbf{I}_T$. As a result, $\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i / \sigma_{\boldsymbol{\eta}, ii}$ follows a chi-square distribution of degree $T - r$. Thus, according to Lemma 2, we can obtain that

$$\begin{aligned} \max_{i \in \mathcal{I}_0} |T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii}| &\leq \max_{i \in \mathcal{I}_0} |T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i - (T - r) \sigma_{\boldsymbol{\eta}, ii} / T| + r \max_{i \in \mathcal{I}_0} \sigma_{\boldsymbol{\eta}, ii} / T \\ &= \max_{i \in \mathcal{I}_0} \sigma_{\boldsymbol{\eta}, ii} |T^{-1} \chi_{T-r} - (T - r) / T| + r \max_i \sigma_{\boldsymbol{\eta}, ii} / T \\ &= O_p(\sqrt{\log N/T}) + O_p(T^{-1}) = O_p(\sqrt{\log N/T}). \end{aligned}$$

Consequently, we can obtain

$$\max_{i \in \mathcal{I}_0} |T^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii}| = O_p(\sqrt{\log N/T}). \quad (\text{A.2})$$

[†]Wei Lan is from Statistics School and Center of Statistical Research, Southwestern University of Finance and Economics (E-mail: facelw@gmail.com); Lilun Du is from the Department of ISOM, Hong Kong University of Science and Technology (E-mail: dulilun@ust.hk).

We next consider the first part of (A.1). Note that $T^{-1}\boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\boldsymbol{\varepsilon}_i = T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\boldsymbol{\eta}_i + 2T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i + T^{-1}\boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i$. We then consider the three parts in the following three steps.

STEP I. We firstly consider $T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\boldsymbol{\eta}_i - T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z})\boldsymbol{\eta}_i$. According to the results of Theorem 2 in Wang (2012) that $\text{tr}\{\mathcal{Q}(\mathbf{Z}) - \mathcal{Q}(\bar{\mathbf{Z}})\}^2 = O_p(T^{-1})$, we can obtain that

$$\begin{aligned} \max_{i \in \mathcal{I}_0} |T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\boldsymbol{\eta}_i - T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{Z})\boldsymbol{\eta}_i| &= \max_{i \in \mathcal{I}_0} |T^{-1}\boldsymbol{\eta}_i^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}\boldsymbol{\eta}_i| \\ &\leq [\text{tr}\{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}^2]^{1/2} \max_{i \in \mathcal{I}_0} T^{-1}\boldsymbol{\eta}_i^\top \boldsymbol{\eta}_i = O_p(T^{-1/2}), \end{aligned}$$

where the last equality is due to the fact that $\max_i \boldsymbol{\eta}_i^\top \boldsymbol{\eta}_i = O_p(T)$ by Lemma 2.

STEP II. We next consider $T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i$, which is equivalent to $T^{-1}\boldsymbol{\eta}_i^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}\mathbf{Z}\boldsymbol{\gamma}_i$. By the results of Theorem 2 in Wang (2012) that $\text{tr}\{\mathcal{Q}(\mathbf{Z}) - \mathcal{Q}(\bar{\mathbf{Z}})\}^2 = O_p(T^{-1})$, together with condition (C4), we have

$$\begin{aligned} &|\boldsymbol{\eta}_i^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}\mathbf{Z}\boldsymbol{\gamma}_i|^2 \\ &= \boldsymbol{\eta}_i^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}\mathbf{Z}\boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}\boldsymbol{\eta}_i \\ &\leq \max_i \|\boldsymbol{\gamma}_i\|^2 \times \lambda_{\max}(\mathbf{Z}\mathbf{Z}^\top) \times \boldsymbol{\eta}_i^\top \{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}^2 \boldsymbol{\eta}_i \\ &\leq T \times \text{tr}\{\mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}^2 \times \|\boldsymbol{\eta}_i\|^2 = O_p(T). \end{aligned}$$

Hence, $T^{-1}\boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i = O_p(T^{-1/2})$.

STEP III. We lastly consider $T^{-1}\boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i$. By condition (C4) and the fact that $\text{tr}(\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}) = O_p(1)$ as demonstrated by Wang (2012), we obtain

$$\max_{i \in \mathcal{I}_0} T^{-1}\boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}\boldsymbol{\gamma}_i \leq T^{-1} \max_{i \in \mathcal{I}_0} \|\boldsymbol{\gamma}_i\|^2 \text{tr}(\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}})\mathbf{Z}) = O_p(1/T) = o_p\left(\sqrt{\log N/T}\right).$$

Combining all of these results above, we have completed the first part of the proof of Lemma 5.

We next consider the second part of Lemma 5. Note that $\mathcal{Q}(\mathbb{X})\mathcal{E}^\top$ can be written as $\mathcal{Q}(\mathbb{X})\mathbf{Z}\boldsymbol{\gamma}^\top + \mathcal{Q}(\mathbb{X})\boldsymbol{\eta}^\top$. As a result, $\mathcal{Q}(\mathbb{X})\mathcal{E}^\top$ also follows a latent factor structure of dimension r . The only difference is that, the variance of the random error is converted to be $(T - r)\sigma_{\boldsymbol{\eta}, ii}/T$. Thus, by the results in the first part of the proof, we can obtain that

$$\max_{i \in \mathcal{I}_0} |\tilde{\sigma}_{\boldsymbol{\eta}, ii} - (T - r)\sigma_{\boldsymbol{\eta}, ii}/T| = O_p\left\{\sqrt{\log N/T}\right\}.$$

Consequently, we have

$$\begin{aligned} \max_{i \in \mathcal{I}_0} |\tilde{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| &\leq \max_i |\tilde{\sigma}_{\boldsymbol{\eta}, ii} - (T - r)\sigma_{\boldsymbol{\eta}, ii}/T| + \max_{i \in \mathcal{I}_0} r(\sigma_{\boldsymbol{\eta}, ii})/T \\ &= O_p\left\{\sqrt{\log N/T}\right\} + O(1/T) = O_p\left\{\sqrt{\log N/T}\right\}, \end{aligned}$$

which completes the entire proof.

Proof of Lemma 6

Proof. Similar to the proof of Lemma 5, we only need to prove that

$$\max_{i \in \mathcal{I}_0} |\boldsymbol{w}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i| = O_p\left\{(\log N)^{1/4}/T^{1/2}\right\} + O_p(T^{-1/4}) + O_p\{N^{-\nu}\}.$$

By triangle inequality inequality, we can obtain

$$\begin{aligned} |\boldsymbol{w}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i| &= |\boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i + \boldsymbol{w}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i| \\ &\leq |\boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i| + |\boldsymbol{w}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i|. \end{aligned} \tag{A.3}$$

We then consider the above two terms separately. Note that

$$\begin{aligned} |\boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i|^2 &= (\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^\top \left(\mathbf{Z}^\top \boldsymbol{w} \boldsymbol{w}^\top \mathbf{Z} \right) (\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i) \\ &\leq \lambda_{\max}(\mathbf{Z}^\top \boldsymbol{w} \boldsymbol{w}^\top \mathbf{Z}) \|\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|^2 = O_p(r) \|\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|^2, \end{aligned}$$

where the last result above is due to the fact that $\boldsymbol{w}^\top \mathbf{Z}$ has mean zero and identity covariance matrix given $\boldsymbol{w}$. As a result,

$$\begin{aligned} \max_{i \in \mathcal{I}_0} |\boldsymbol{w}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i - \boldsymbol{w}^\top \mathbf{Z} \boldsymbol{\gamma}_i| &\leq \max_{i \in \mathcal{I}_0} \|\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\| \\ &= \max_{i \in \mathcal{I}_0} \|T^{-1} \bar{\mathbf{Z}}^\top \mathbf{Z} \boldsymbol{\gamma}_i + T^{-1} \bar{\mathbf{Z}}^\top \boldsymbol{\eta}_i - \boldsymbol{\gamma}_i\| \leq T^{-1} \max_{i \in \mathcal{I}_0} \|\bar{\mathbf{Z}}^\top \mathbf{Z} \boldsymbol{\gamma}_i - \mathbf{Z}^\top \mathbf{Z} \boldsymbol{\gamma}_i\| \\ &\quad + T^{-1} \max_{i \in \mathcal{I}_0} \|\bar{\mathbf{Z}}^\top \boldsymbol{\eta}_i - \bar{\mathbf{Z}}^\top \bar{\boldsymbol{\eta}}_i\|. \end{aligned} \tag{A.4}$$

We consider the above two parts separately. First note that

$$\begin{aligned} \|\bar{\mathbf{Z}}^\top \mathbf{Z} \boldsymbol{\gamma}_i - \mathbf{Z}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|^2 &= \boldsymbol{\gamma}_i^\top \mathbf{Z}^\top (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathbf{Z} \boldsymbol{\gamma}_i \\ &\leq \|\boldsymbol{\gamma}_i\|^2 \text{tr}\left\{ \mathbf{Z}^\top (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathbf{Z} \right\} \leq T \|\boldsymbol{\gamma}_i\|^2 \text{tr}\left\{ (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \right\}. \end{aligned}$$

Moreover, according to condition (C4), we have $\max_i \|\gamma_i\|^2 = O(1)$. Consequently, we only need to consider $\text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\}$. According to condition (C5), one can easily verify that

$$\text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\} = \sum_{e \leq r} \|\bar{\boldsymbol{\epsilon}}_e - \boldsymbol{\epsilon}_e\|^2 = O_p(N^{-2\nu}).$$

Here, $(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_r) = \mathbf{Z}$. Consequently, we have

$$T^{-1} \max_{i \in \mathcal{I}_0} \|\bar{\mathbf{Z}}^\top \mathbf{Z} \gamma_i - \mathbf{Z}^\top \mathbf{Z} \gamma_i\| = O_p(N^{-\nu}/\sqrt{T}). \quad (\text{A.5})$$

We next consider the second part of (A.4). By Cauchy-Schwarz inequality, we have

$$\|\bar{\mathbf{Z}}^\top \bar{\boldsymbol{\eta}}_i - \bar{\mathbf{Z}}^\top \boldsymbol{\eta}_i\|^2 \leq 2\|(\bar{\mathbf{Z}} - \mathbf{Z})^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2 + 2\|\mathbf{Z}^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2.$$

We firstly consider the second term. Note that $\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i = \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z} \gamma_i - \mathcal{H}(\bar{\mathbf{Z}}) \boldsymbol{\eta}_i$, where $\mathcal{H}(\cdot) = \mathbf{I}_T - \mathcal{Q}(\cdot)$. Consequently, applying Cauchy-Schwarz inequality, we have $\|\mathbf{Z}^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2 \leq 2\gamma_i^\top \{\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z}\}^2 \gamma_i + 2\boldsymbol{\eta}_i^\top \mathcal{H}(\bar{\mathbf{Z}}) \mathbf{Z} \mathbf{Z}^\top \mathcal{H}(\bar{\mathbf{Z}}) \boldsymbol{\eta}_i \leq 2\|\gamma_i\|^2 \text{tr}\{\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z}\}^2 + 2\text{tr}\{\mathbf{Z}^\top \mathcal{H}(\bar{\mathbf{Z}}) \mathbf{Z}\} \boldsymbol{\eta}_i^\top \mathcal{H}(\bar{\mathbf{Z}}) \boldsymbol{\eta}_i$. The first term is again $O_p(1)$ by condition (C4) and the results from Wang (2012) that $\text{tr}\{\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z}\} = O_p(1)$. For the second term, note that $\mathcal{H}(\bar{\mathbf{Z}})$ is a projection matrix of rank r , then $\boldsymbol{\eta}_i^\top \mathcal{H}(\bar{\mathbf{Z}}) \boldsymbol{\eta}_i$ can be expressed as a summation of r chi-square distributions of degree 1. Accordingly, by Bonferroni inequality, we have $\max_{i \in \mathcal{I}_0} \boldsymbol{\eta}_i^\top \mathcal{H}(\bar{\mathbf{Z}}) \boldsymbol{\eta}_i = O_p(\sqrt{\log N})$. This together with the results in Theorem 2 of Wang (2012) that $\text{tr}\{\mathbf{Z}^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z}\} = O_p(1)$ yields that $\|\mathbf{Z}^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2 = O_p(\sqrt{\log N})$. We next consider the second part,

$$\begin{aligned} \|(\bar{\mathbf{Z}} - \mathbf{Z})^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2 &= (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)^\top (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i) \\ &\leq \text{tr}\{(\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top\} \|\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i\|^2 \leq \text{tr}\{(\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top\} \max_{i \in \mathcal{I}_0} \|\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i\|^2. \end{aligned}$$

Similar to the proof of Lemma 5, we can have $\max_{i \in \mathcal{I}_0} \|\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i\|^2 = O_p(\sqrt{\log N})$. Moreover, $\text{tr}\{(\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top\} = O_p(N^{-2\nu})$ according to condition (C5). As a result, $\|(\bar{\mathbf{Z}} - \mathbf{Z})^\top (\bar{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)\|^2 = O_p\{\sqrt{\log N}/N^{2\nu}\}$. Consequently, combining these results above, we have

$$T^{-1} \max_{i \in \mathcal{I}_0} \|\bar{\mathbf{Z}}^\top \boldsymbol{\eta}_i - \bar{\mathbf{Z}}^\top \bar{\boldsymbol{\eta}}_i\| = O_p(\{\log N\}^{1/4}/T^{1/2}). \quad (\text{A.6})$$

Combining these results in (A.5) and (A.6), together with condition (C1),

$$\max_{i \in \mathcal{I}_0} |\mathbf{w}^\top \mathbf{Z} \bar{\gamma}_i - \mathbf{w}^\top \mathbf{Z} \gamma_i| = O_p\left(\{\log N\}^{1/4}/T^{1/2}\right) + O_p\left(N^{-\nu}/\sqrt{T}\right). \quad (\text{A.7})$$

We next consider the second term of (A.3). Note that

$$\|\mathbf{w}^\top \bar{\mathbf{Z}} \bar{\gamma}_i - \mathbf{w}^\top \mathbf{Z}^\top \bar{\gamma}_i\|^2 = \bar{\gamma}_i^\top (\mathbf{Z} - \bar{\mathbf{Z}}) \mathbf{w} \mathbf{w}^\top (\mathbf{Z} - \bar{\mathbf{Z}})^\top \bar{\gamma}_i \leq \|\bar{\gamma}_i\|^2 \text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\}.$$

As a result, we can obtain that

$$\begin{aligned} \max_{i \in \mathcal{I}_0} |\mathbf{w}^\top \bar{\mathbf{Z}}^\top \bar{\gamma}_i - \mathbf{w}^\top \mathbf{Z}^\top \bar{\gamma}_i| &\leq O_p(N^{-\nu}) \max_{i \in \mathcal{I}_0} \|\bar{\gamma}_i\|. \\ &\leq O_p(N^{-\nu}) \max_{i \in \mathcal{I}_0} \left\{ \|\gamma_i\| + \|\bar{\gamma}_i - \gamma_i\| \right\} = O_p(N^{-\nu}). \end{aligned} \quad (\text{A.8})$$

Combining these results in (A.7) and (A.8), we thus have

$$\max_{i \in \mathcal{I}_0} \|\mathbf{w}^\top \bar{\mathbf{Z}}^\top \bar{\gamma}_i - \mathbf{w}^\top \mathbf{Z}^\top \gamma_i\| = O_p\left(\{\log N\}^{1/4}/T^{1/2}\right) + O_p\left(N^{-\nu}\right),$$

which completes the entire proof of Lemma 6.

Proof of Proposition 1

Proof. We first prove that $V^\circ(t)/N_0$ converges to the uniform distribution t almost surely. Since $\boldsymbol{\eta}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma_{\boldsymbol{\eta}, ii} \mathbf{I}_T$, T_i° is a standard normal variable under the null hypothesis. Hence, it amounts to show that

$$N_0^{-1} \sum_{i \in \mathcal{I}_0} \left[\mathbf{I}(T_i^\circ \geq -z_{t/2}) - \mathbf{P}(T_i^\circ \geq -z_{t/2}) \right] \xrightarrow{N_0 \rightarrow \infty} 0 \quad a.s. \quad (\text{A.9})$$

$$N_0^{-1} \sum_{i \in \mathcal{I}_0} \left[\mathbf{I}(T_i^\circ \leq z_{t/2}) - \mathbf{P}(T_i^\circ \leq z_{t/2}) \right] \xrightarrow{N_0 \rightarrow \infty} 0 \quad a.s. \quad (\text{A.10})$$

By Lemma 1, the conclusion (A.9) is valid if we can show that

$$\text{var} \left\{ N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^\circ \geq -z_{t/2}) \right\} = O(N_0^{-\delta}), \quad \text{for some } \delta > 0. \quad (\text{A.11})$$

Starting with (A.11), we note that

$$\begin{aligned}
& \text{var} \left\{ N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o \geq -z_{t/2}) \right\} \\
&= N_0^{-2} \sum_{i \in \mathcal{I}_0} \text{var} \left\{ \mathbf{I}(T_i^o \geq -z_{t/2}) \right\} \\
&+ N_0^{-2} \sum_{i_1 \in \mathcal{I}_0, i_2 \in \mathcal{I}_0, i_1 \neq i_2} \text{cov} \left\{ \mathbf{I}(T_{i_1}^o \geq -z_{t/2}), \mathbf{I}(T_{i_2}^o \geq -z_{t/2}) \right\}.
\end{aligned}$$

The first term above is $O_p(N_0^{-1})$ by the fact that $\text{var} \{ \mathbf{I}(T_i^o \geq -z_{t/2}) \} \leq 1$. For the second term, the covariance is given by

$$\begin{aligned}
& \mathbf{P}(T_{i_1}^o \geq -z_{t/2}, T_{i_2}^o \geq -z_{t/2}) - \mathbf{P}(T_{i_1}^o \geq -z_{t/2}) \mathbf{P}(T_{i_2}^o \geq -z_{t/2}) \\
&= \mathbf{P}(T_{i_1}^o \geq -z_{t/2}, T_{i_2}^o \geq -z_{t/2}) - \{ \Phi(z_{t/2}) \}^2.
\end{aligned} \tag{A.12}$$

To evaluate (A.12), we need to verify that for any $i_1 \in \mathcal{I}_0$ and $i_2 \in \mathcal{I}_0$, $(T_{i_1}^o, T_{i_2}^o)^\top$ has a bivariate normal distribution with mean zero and covariance matrix given by $\Sigma_{i_1 i_2} = (\sigma_{i_1 i_2})_{2 \times 2}$, where $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{12} = \sigma_{21} = \rho_{\boldsymbol{\eta}, i_1 i_2}$. This conclusion is immediately implied by the assumption that $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_T\}$ are independent and the constraint that $\mathbf{w}^\top \mathbf{w} = 1$. Without loss of generality, we assume $\rho_{\boldsymbol{\eta}, i_1 i_2} > 0$ (for $\rho_{\boldsymbol{\eta}, i_1 i_2} < 0$, the conclusion is similar). Let (x, y, z) be three independent standard normal random variables. Then $T_{i_1}^o, T_{i_2}^o$ can be constructed by $T_{i_1}^o = (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z + (1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} x$, $T_{i_2}^o = (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z + (1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} y$. By using the above formulas, the first term in (A.12) can be expressed as

$$\begin{aligned}
& \mathbf{P}(T_{i_1}^o \geq -z_{t/2}, T_{i_2}^o \geq -z_{t/2}) \\
&= \mathbf{P} \left((\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z + (1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} x \geq -z_{t/2}, (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z + (1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} y \geq -z_{t/2} \right) \\
&= \mathbf{P} \left(x \geq \frac{-z_{t/2} - (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z}{(1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2}}, y \geq \frac{-z_{t/2} - (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z}{(1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2}} \right) \\
&= \int_{-\infty}^{\infty} \Phi \left\{ \frac{(\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} + z_{t/2}}{(1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2}} \right\}^2 \phi(z) dz.
\end{aligned} \tag{A.13}$$

Employing Taylor expansion to $\Phi(\cdot)$ with respect to $(\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2}$ yields that

$$\Phi \left\{ \frac{(\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} z + z_{t/2}}{(1 - \rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2}} \right\} = \Phi(z_{t/2}) + \phi(z_{t/2}) z (\rho_{\boldsymbol{\eta}, i_1 i_2})^{1/2} + \frac{1}{2} \phi(z_{t/2}) (1 - z^2) \rho_{\boldsymbol{\eta}, i_1 i_2} + R(\rho_{\boldsymbol{\eta}, i_1 i_2}),$$

where $R(\rho_{\boldsymbol{\eta}, i_1 i_2}) = f_R(z) O(|\rho_{\boldsymbol{\eta}, i_1 i_2}|^{3/2})$ with $f_R(z)$ being a polynomial function of z of

order 3. Therefore, using the fact that $E(z) = 0, \text{var}(z) = 1$, (A.13) equals

$$\Phi(z_{t/2})^2 + \phi(z_{t/2})^2 \rho_{\boldsymbol{\eta}, i_1 i_2} + O(|\rho_{\boldsymbol{\eta}, i_1 i_2}|^{3/2}). \quad (\text{A.14})$$

Combining (A.12) and (A.14), we obtain that

$$\left| \text{cov} \left\{ \mathbf{I}(T_{i_1}^o \geq -z_{t/2}), \mathbf{I}(T_{i_2}^o \geq -z_{t/2}) \right\} \right| \leq \phi(z_{t/2})^2 \rho_{\boldsymbol{\eta}, i_1 i_2} + O(|\rho_{\boldsymbol{\eta}, i_1 i_2}|^{3/2}).$$

By the assumption (2.3) and $\pi_0 > 0$, $N_0^{-2} \sum_{i_1 \in \mathcal{I}_0, i_2 \in \mathcal{I}_0, i_1 \neq i_2} |\rho_{\boldsymbol{\eta}, i_1 i_2}| = O(N_0^{-\delta})$. This in turn implies that

$$\text{var} \left\{ N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o \geq -z_{t/2}) \right\} = O_p(N_0^{-\delta}), \quad \text{for some } \delta > 0,$$

from which (A.9) holds. (A.10) can be verified in a similar manner.

Now we show that $V(t)/N_0$ also converges to the uniform distribution t almost surely under the conditions that $\max_{i \in \mathcal{I}_0} |\widehat{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| \rightarrow_p 0$ and $\max_{i \in \mathcal{I}_0} |(\mathbf{1}^\top \mathbf{Q} \mathbf{1})^{-1/2} \mathbf{1}^\top \mathbf{Q} \widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i - (\mathbf{1}^\top \mathbf{Q} \mathbf{1})^{-1/2} \mathbf{1}^\top \mathbf{Q} \mathbf{Z} \boldsymbol{\gamma}_i| \rightarrow_p 0$. Denote by $t_1 = -z_{t/2}(\widehat{\sigma}_{\boldsymbol{\eta}, ii})^{1/2}/(\sigma_{\boldsymbol{\eta}, ii})^{1/2} + (\mathbf{1}^\top \mathbf{Q} \mathbf{1})^{-1/2} \mathbf{1}^\top \mathbf{Q}(\widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i - \mathbf{Z} \boldsymbol{\gamma}_i)/(\sigma_{\boldsymbol{\eta}, ii})^{1/2}$ and $t_2 = z_{t/2}(\widehat{\sigma}_{\boldsymbol{\eta}, ii})^{1/2}/(\sigma_{\boldsymbol{\eta}, ii})^{1/2} + (\mathbf{1}^\top \mathbf{Q} \mathbf{1})^{-1/2} \mathbf{1}^\top \mathbf{Q}(\widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i - \mathbf{Z} \boldsymbol{\gamma}_i)/(\sigma_{\boldsymbol{\eta}, ii})^{1/2}$. Then, by the assumptions of Proposition 1, we have $t_1 \rightarrow_p -z_{t/2}$ and $t_2 \rightarrow_p z_{t/2}$ uniformly for any i . For any $\varepsilon > 0$, there exists a constant $\delta > 0$ not depending on i , such that $\mathbf{P}(|t_1 + z_{t/2}| < \delta) \geq 1 - \varepsilon$ and $\mathbf{P}(|t_2 - z_{t/2}| < \delta) \geq 1 - \varepsilon$. Let $\Omega_1 = \{w : |t_1(w) + z_{t/2}| < \delta, |t_2(w) - z_{t/2}| < \delta\}$. As a result, $\mathbf{P}(\Omega_1) \geq 1 - \varepsilon$. Denote by Ω_2 the set of w such that $V^o(t)/N_0$ converges to t such that $\mathbf{P}(\Omega_2) = 1$. For any $w \in \Omega_1 \cap \Omega_2$, we bound $V(t)/N_0$ as

$$\begin{aligned} N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) > -z_{t/2} + \delta) &\leq N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) > t_1(w)) \leq N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) > -z_{t/2} - \delta) \\ N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) < z_{t/2} - \delta) &\leq N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) < t_2(w)) \leq N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) < z_{t/2} + \delta) \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ and $\delta \rightarrow 0$, we obtain that for all $w \in \Omega_1 \cap \Omega_2$ with $\mathbf{P}(\Omega_1 \cap \Omega_2) = 1$, $N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) > t_1(w)) \rightarrow t/2$ and $N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{I}(T_i^o(w) < t_2(w)) \rightarrow t/2$. This completes the proof of Proposition 1.

Proof of Theorem 1

Proof. By Proposition 1 and the sparsity condition (C2), $R^o(t)/N \geq V^o(t)/N$ which is lower bounded by $t/2$ as N, T are sufficiently large. By the result of Proposition 1,

$V(t)/N_0 - V^o(t)/N_0 \xrightarrow{a.s.} 0$, which implies that $R(t)/N$ is lower bounded by $t/2$. Then, for any given $t > 0$,

$$\begin{aligned}
& \lim_{N,T \rightarrow \infty} |\widehat{\text{FDR}}_\lambda(t) - \text{FDP}^o(t)| \\
& \leq \lim_{N,T \rightarrow \infty} \frac{\{\widehat{\pi}_0(\lambda)t\}\{R^o(t)/N\} - \{V^o(t)/N\}\{R(t)/N\}}{\{R^o(t)/N\}\{R(t)/N\}} \\
& \leq 4/t^2 \times \lim_{N,T \rightarrow \infty} \left\{ |\widehat{\pi}_0(\lambda)t - V^o(t)/N| \{R^o(t)/N\} + \{V^o(t)/N\} |R^o(t)/N - R(t)/N| \right\} \\
& \leq 4/t^2 \times \left\{ \lim_{N,T \rightarrow \infty} |\widehat{\pi}_0(\lambda) - \pi_0|t + |\pi_0t - V^o(t)/N| + |R(t)/N - R^o(t)/N| \right\} \\
& \leq 4/t^2 \times \left\{ \lim_{N,T \rightarrow \infty} |\widehat{\pi}_0(\lambda) - \pi_0|t + |\pi_0t - V^o(t)/N| + |V(t)/N - V^o(t)/N| + O(N_1/N) \right\}.
\end{aligned}$$

According to the proof of Propositions 1 and 2, $|V(t)/N - V^o(t)/N| \leq C \times \{\max_{i \in \mathcal{I}_0} |\widehat{\sigma}_{\eta,ii} - \sigma_{\eta,ii}| + \max_{i \in \mathcal{I}_0} |(\mathbf{1}^\top \mathbf{Q}(\mathbb{X})\mathbf{1})^{-1/2} \mathbf{1}^\top \mathbf{Q}(\mathbb{X})(\widehat{\mathbf{Z}}\widehat{\boldsymbol{\gamma}}_i - \mathbf{Z}\boldsymbol{\gamma}_i)|\} = O_p\{(\log N)^{1/4}/T^{1/2}\} + O_p(N^{-\nu}) + O_p(N^{-1}T\|\boldsymbol{\mu}\|^2) + O_p(N_1^{1/2}N^{-1}T\|\boldsymbol{\mu}\|)$, and $|\pi_0t - V^o(t)/N| = O_p(N^{-\delta/2})$. Combining these two results, $|V(t)/N_0 - t| = O_p\{(\log N)^{1/4}/T^{1/2}\} + O_p(N^{-\nu}) + O_p(N^{-1}T\|\boldsymbol{\mu}\|^2) + O_p(N_1^{1/2}N^{-1}T\|\boldsymbol{\mu}\|) + O_p(N^{-\delta/2})$. By decomposition, $\widehat{\pi}_0(\lambda) = \sum_{i \in \mathcal{I}_0} \mathbf{I}(P_i > \lambda) / \{N(1 - \lambda)\} + O(N_1/N)$. Similar to the proof of Proposition 1, we can show that $\sum_{i \in \mathcal{I}_0} \mathbf{I}(P_i > \lambda) / \{N(1 - \lambda)\} \rightarrow \pi_0$ almost surely. Moreover, the convergence rate of $\sum_{i \in \mathcal{I}_0} \mathbf{I}(P_i > \lambda) / \{N(1 - \lambda)\}$ is the same as that of $V(t)/N_0 - t$, which completes the proof of Theorem 1.

The “negative” dependence of BH type estimator in finite sample

It has been theoretically shown that the BH type estimator can control the FDR when the test statistics or p -values have some special dependence structures (Benjamini and Yekutieli, 2001; Storey et al., 2004). In this part, we will show that the BH type estimator is negatively correlated with the true FDP when the signals under the alternative are strong in finite sample.

For ease of presentation, denote by $P_1, \dots, P_N$ are p -values corresponding to N hypotheses $H_i, i = 1, \dots, N$, and there are N_0 of them are from the true null. Then, $V(t)$, $R(t)$, and $\text{FDP}(t)$ can be defined in a similar way as in (2.10) in the main article. Benjamini and Hochberg (1995) proposed to estimate the $\text{FDP}(t)$ as $\widehat{\text{FDP}}(t) = Nt / \max(R(t), 1)$. If the p -values under the alternative are all smaller than the threshold t , then $R(t) = N_1 + V(t)$, where t is a threshold and $N_1 = N - N_0$. For any two replications, the number of the false discoveries, the true discoveries, and the false

discovery proportion are denoted as $V_i(t)$, $R_i(t)$, and $\text{FDP}_i(t)$ for $i = 1, 2$, respectively. Then, the slope of the two pairs of $\{\text{FDP}_i(t), \widehat{\text{FDP}}_i(t)\}$ which satisfy the above property and $V_1(t) \neq V_2(t)$, can be explicitly decomposed as,

$$\begin{aligned} \text{slope}_{\text{BH}} &= \frac{\widehat{\text{FDP}}_1(t) - \widehat{\text{FDP}}_2(t)}{\text{FDP}_1(t) - \text{FDP}_2(t)} \\ &= \frac{\frac{Nt}{R_1(t)} - \frac{Nt}{R_2(t)}}{\frac{V_1(t)}{R_1(t)} - \frac{V_2(t)}{R_2(t)}} \\ &= \frac{\frac{Nt}{N_1+V_1(t)} - \frac{Nt}{N_1+V_2(t)}}{\frac{V_1(t)}{N_1+V_1(t)} - \frac{V_2(t)}{N_1+V_2(t)}} \\ &= -\frac{Nt}{N_1} < 0 \end{aligned}$$

However, under the independence or weak dependence assumption of the p -values, it is straightforward to show that $|\widehat{\text{FDP}}(t) - \text{FDP}(t)| = o_p(1)$ when N is sufficient large and $\pi_0 = 1$.

The above negative dependence property can be extended to Storey's estimator (Storey et al., 2004), which incorporates the estimator of $\pi_0 = \lim_{N \rightarrow \infty} N_0/N$ into the BH method. Specifically, the estimated FDP can be formulated as $\frac{N\widehat{\pi}_0(\lambda)t}{R(t)\sqrt{1}}$, with $\widehat{\pi}_0(\lambda) = \#\{P_i; P_i > \lambda\}/\{N(1 - \lambda)\}$ for a tuning parameter $\lambda > 0$. Under the same assumption on the alternative and $t \leq \lambda$, we have $\widehat{\pi}_0(\lambda) = (N_0 - V(\lambda))/\{N(1 - \lambda)\}$. Similarly, the slope of any two pairs of estimators can be derived as

$$\begin{aligned} \text{slope}_{\text{Storey}} &= \frac{t}{1 - \lambda} \times \left\{ -\frac{N_0}{N_1} - \frac{V_1(\lambda) - V_2(\lambda)}{V_1(t) - V_2(t)} - \frac{V_1(\lambda)V_2(t) - V_1(t)V_2(\lambda)}{N_1(V_1(t) - V_2(t))} \right\} \\ &= \frac{t}{1 - \lambda} \times \left\{ -\frac{N_0 - V_2(\lambda)}{N_1} - \left(1 + \frac{V_2(t)}{N_1}\right) \times \frac{V_1(\lambda) - V_2(\lambda)}{V_1(t) - V_2(t)} \right\}, \end{aligned}$$

which is negative if $V_1(\lambda) - V_2(\lambda)$ and $V_1(t) - V_2(t)$ have the same sign.

REFERENCES

- Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 57, 289–300.
- Benjamini, Y. and Yekutieli, D. (2001). "The control of false discovery rate in multiple testing under dependency," *Annals of Statistics*, 29, 1165–1188.

- Fan, J., Han, X., and Gu, W. (2012). “Estimating False Discovery Proportion Under Arbitrary Covariance Dependence,” *Journal of the American Statistical Association*, 107, 1019–1035.
- Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). “Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach,” *Journal of the Royal Statistical Society, Series B*, 66, 187–205.
- Wang, H. (2012). “Factor profiled independence screening,” *Biometrika*, 99, 15–28.