Functional Two-Sample Test based on Projection

Yang Bai[1], Caihong Qin[1] and Huichen Zhu[2]

[1] *Shanghai University of Finance and Economics and*

[2] *The Chinese University of Hong Kong*

Supplementary Material

This Supplementary Material contains a theoretical example in Section S1 to elucidate the conclusion presented in Section 2.2, additional simulations in Section S2 and the technical proofs in Section S3.

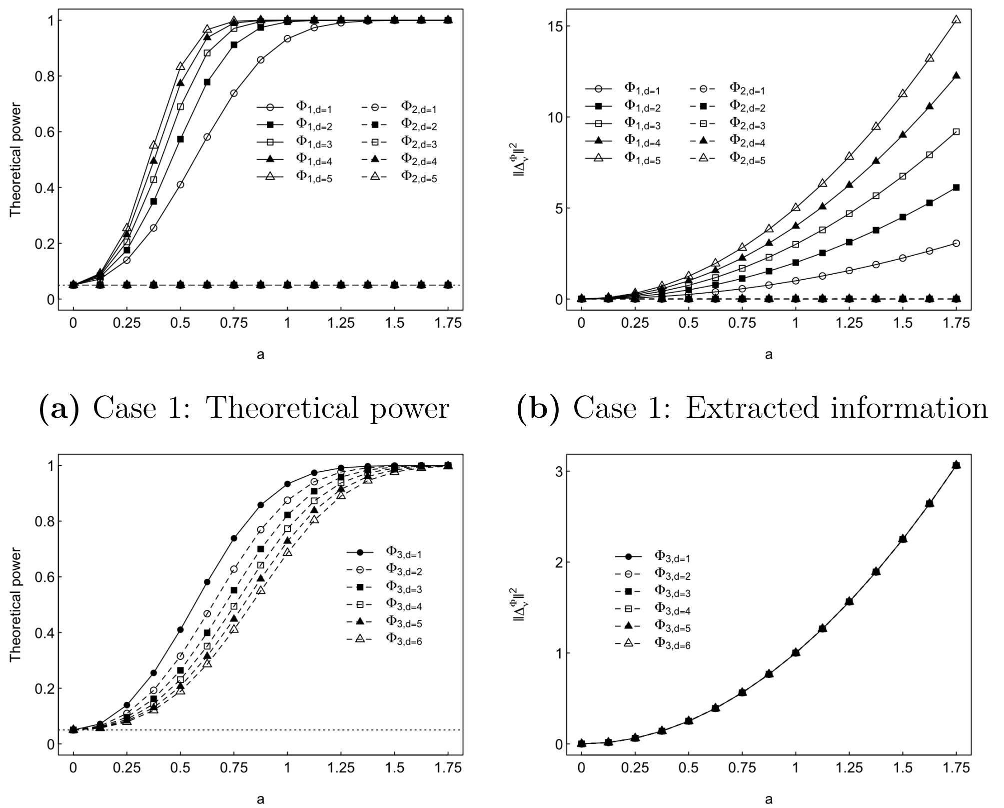
S1 A theoretical example

We present a theoretical example to illustrate how the power function $\beta_{H^\Phi}(\cdot)$ changes with the extracted information $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ and the dimension of the projective space d. The functional observations $\mathbf{y}_{i1}(t),\ldots,\mathbf{y}_{in_i}(t)$ are generated from Gaussian process $\mathrm{GP}(\boldsymbol{\mu}_i(t),\mathcal{K}(s,t))$ for $i=1,2$ with $\boldsymbol{\mu}_1(t)=0$ and $\boldsymbol{\mu}_2(t)=\sum_{k=1}^{K}a_k\psi_k(t)$. We set $a_k=a$ for $k=1,\ldots,K$, $a_k=0$ for $k=K+1,\ldots,K+d$, and $\psi_k(t)=\sqrt{2}\cos(k\pi t)$ for $k=1,\ldots,K+d$. The covariance generator is generated by $\mathcal{K}(s,t)=1$ if $s=t$ and 0 otherwise for simplicity. We let $K=5$, $\alpha=0.05$ and $n_1=n_2=25$. In **Case 1**, we compare the power between two different values of the extracted information $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ with the dimension of the projective space d fixed. In **Case 2**, we investigate the change in power with d when the value of $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ is fixed. The details of the two cases are as follows.

Case 1: In this case we consider

$$\Phi_{1,d} = (\psi_1, \ldots, \psi_d)^{\mathrm{T}} \text{ and } \Phi_{2,d} = (\psi_{K+1}, \ldots, \psi_{K+d})^{\mathrm{T}} \text{ with } d = 1, \ldots, K.$$

Thus, the extracted information based on $\Phi_{1,d}$ and $\Phi_{2,d}$ is $\|\Delta_{\boldsymbol{\nu}}^{\Phi_{1,d}}\|^2 = \sum_{k=1}^{d} a_k^2 = da^2$ and $\|\Delta_{\boldsymbol{\nu}}^{\Phi_{2,d}}\|^2 = 0$ for $d = 1, \ldots, K$, respectively. When the extracted information is equal to 0, the value of the power function is exactly the significance level α. Figure S1.1(a) shows that the power of the test corresponding to a non-zero extracted information is much higher than the power corresponding to a zero extracted information with a fixed d. It also demonstrates that the power increases with d when the projective space is determined by $\Phi_{1,d}$. However, the extracted information changes as well in this case. Figure S1.1(b) further illustrates that the extracted information based on $\Phi_{1,d}$ increases with d. The relationship between power and d when α, n_1, n_2, and $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ are fixed is still unknown in this case.



(a) Case 1: Theoretical power **(b)** Case 1: Extracted information

(c) Case 2: Theoretical power **(d)** Case 2: Extracted information

Figure S1.1: (a) and (b) respectively represent the theoretical power $\beta_{H^\Phi}(\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2; d, n_1, n_2)$ and extracted information $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2$ in Case 1 with $\Phi_{1,d}$ and $\Phi_{2,d}$. Similarly, (c) and (d) respectively show the theoretical power and extracted information in Case 2 with $\Phi_{3,d}$.

Case 2: In this case we consider

$$\Phi_{3,d=1} = (\psi_1)^{\mathrm{T}} \text{ and } \Phi_{3,d} = (\psi_1, \psi_{K+1}, \ldots, \psi_{K+d-1})^{\mathrm{T}} \text{ for } d = 2, \ldots, K+1.$$

The extracted information based on $\Phi_{3,d}$ is $\|\Delta_{\boldsymbol{\nu}}^{\Phi_{3,d}}\|^2 = a_1^2 = a^2$ for $d = 1, \ldots, K+1$. Figure S1.1(c) demonstrates that, with the same value of the extracted information but based on different projective space (as shown in Figure S1.1(d)), the power decreases as the dimension of the projective space d increases.

S2 Additional simulation results

In this section, we illustrate the finite sample behavior of the projection tests based on different projection functions discussed in Section 3 with different d. The data generation process is the same as in Section 4. In each of 1000 repetitions we generate $n_1 = n_2 = 150$ observations.

Table S2.1 presents the change of size (when $a = 0$) of the tests for $d = 1, \ldots, 15$. We observe that BS is close to 0.05 when d is small but increases above 0.05 as d increases. This is due to overfitting in (2.7) when d becomes too large, resulting in a higher probability of falsely rejecting the null hypothesis. The solution of CG tends to be stable as d increases.

Table S2.1: The empirical size of the tests based on different projection functions with $a = 0$, $n_1 = n_2 = 150$ and $d = 1, \ldots, 15$

d	BS	CG	PC_K_λ	PC_K_Q	PC_M_λ	PC_M_Q
1	0.064	0.016	0.052	0.156	0.052	0.164
2	0.060	0.064	0.040	0.104	0.040	0.184
3	0.048	0.064	0.048	0.096	0.052	0.164
4	0.064	0.088	0.052	0.072	0.056	0.124
5	0.048	0.064	0.044	0.060	0.052	0.092
6	0.048	0.064	0.052	0.052	0.060	0.068
7	0.056	0.064	0.052	0.052	0.048	0.048
8	0.068	0.056	0.044	0.044	0.040	0.040
9	0.068	0.064	0.052	0.052	0.036	0.036
10	0.068	0.064	0.052	0.052	0.020	0.020
11	0.072	0.056	0.040	0.040	0.020	0.020
12	0.076	0.056	0.048	0.048	0.012	0.012
13	0.068	0.056	0.040	0.040	0.008	0.008
14	0.080	0.056	0.052	0.052	0.004	0.004
15	0.072	0.056	0.048	0.048	0.004	0.004

PC_K_λ remains stable and close to 0.05 for different d, while PC_M_λ stays around 0.05 only when d is not too large. This is because $\widehat{\mathcal{M}}$ contains the sample between-population covariation, which increases the variance in the projected data and interferes with the mean test as d increases. Both PC_K_Q and PC_M_Q are significantly greater than 0.05 when d is small. However, as d increases, they have the same subset as PC_K_λ and PC_M_λ, respectively, resulting in similar performance. Although PC_K_Q and PC_M_Q are unstable with given d, we will demonstrate that they perform well when using the data-driven d selected by the method in Section 3.2 in the following section.

S3 All technique proofs

Proof of Theorem 1. Under H_0, i.e. $\boldsymbol{\mu}=0$, $\boldsymbol{\nu}=\Phi\boldsymbol{\mu}=0$ holds, i.e. H_0^{Φ} holds. Under H_1, i.e. $\boldsymbol{\mu}\neq 0$, by assumption that $\Phi\boldsymbol{\mu}\neq 0$ if $\boldsymbol{\mu}\neq 0$, H_1^{Φ} holds. Thus testing H_0 is equivalent to test H_0^{Φ}. What's more,

$$\beta_{H_0}(\cdot)=pr(I_{reject}\mid H_0)=pr(I_{reject}\mid H_0^{\Phi})=\beta_{H_0^{\Phi}}(\cdot) \tag{S3.1}$$

$$\beta_{H_1}(\cdot)=pr(I_{reject}\mid H_1)=pr(I_{reject}\mid H_1^{\Phi})=\beta_{H_1^{\Phi}}(\cdot) \tag{S3.2}$$

that is, $\beta_H(\cdot)=\beta_{H^{\Phi}}(\cdot)$. □

Proof of Proposition 1. $\Phi_d=(\phi_1,\ldots,\phi_d)^{\mathrm{T}}$ is a multivariate function such that $\phi_k(k=1,2,\ldots)$ generate the range of $\mathcal{K}$. Record $\mathcal{K}_d^{-1}=\Phi^{\mathrm{T}}(\Phi\mathcal{K}\Phi^{\mathrm{T}})^{-1}\Phi$, then $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2=\boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}_d^{-1}\boldsymbol{\mu}$. By monotone convergence theorem, $\boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}_d^{-1}\boldsymbol{\mu}$ converges to $\boldsymbol{\mu}^{\mathrm{T}}\mathcal{K}^{-1}\boldsymbol{\mu}$ as $d\to\infty$, i.e. $\|\Delta_{\boldsymbol{\nu}}^{\Phi}\|^2\overset{d\to\infty}{\to}\|\Delta_{\boldsymbol{\mu}}\|^2$. □

Proof of Theorem 2. We write

$$|\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\phi}_t^{\mathrm{CG}}-\boldsymbol{\mu}^{\mathrm{T}}\phi_t^{\mathrm{CG}}|\leq\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|\|\widehat{\phi}_t^{\mathrm{CG}}\|+|\boldsymbol{\mu}^{\mathrm{T}}(\widehat{\phi}_t^{\mathrm{CG}}-\phi_t^{\mathrm{CG}})|,$$

$$|(\widehat{\phi}_t^{\mathrm{CG}})^{\mathrm{T}}\widehat{\mathcal{K}}\widehat{\phi}_t^{\mathrm{CG}}-(\phi_t^{\mathrm{CG}})^{\mathrm{T}}\mathcal{K}\phi_t^{\mathrm{CG}}|\leq\|\widehat{\phi}_t^{\mathrm{CG}}\|\|\widehat{\mathcal{K}}-\mathcal{K}\|_{\infty}\|\widehat{\phi}_t^{\mathrm{CG}}\|+|(\widehat{\phi}_t^{\mathrm{CG}})^{\mathrm{T}}\mathcal{K}\widehat{\phi}_t^{\mathrm{CG}}-(\phi_t^{\mathrm{CG}})^{\mathrm{T}}\mathcal{K}\phi_t^{\mathrm{CG}}|.$$

Proceeding as in the proof of Theorem 1 in Kraus and Stefanucci (2019), we can show that

$$|\boldsymbol{\mu}^{\mathrm{T}}(\widehat{\phi}_t^{\mathrm{CG}} - \phi_t^{\mathrm{CG}})| = O_p(n_0^{-1/2}\omega_t^{-1}\|\gamma^{(d)}\| + n_0^{-1}\omega_t^{-3}),$$
$$|(\widehat{\phi}_t^{\mathrm{CG}})^{\mathrm{T}}\mathcal{K}\widehat{\phi}_t^{\mathrm{CG}} - (\phi_t^{\mathrm{CG}})^{\mathrm{T}}\mathcal{K}\phi_t^{\mathrm{CG}}| = O_p(n_0^{-1/2}\omega_t^{-1}\|\gamma^{(d)}\| + n_0^{-1}\omega_t^{-3}).$$

Combining this with the facts that $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = O_p(n_0^{-1/2})$, $\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty = O_p(n_0^{-1/2})$ and $\|\widehat{\phi}_t^{\mathrm{CG}}\| = O_p(1)$ gives

$$\|\widehat{\Delta_{\boldsymbol{\nu}}^{\mathrm{CG},t}}\|^2 = \frac{(\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\phi}_t^{\mathrm{CG}})^2}{\langle\widehat{\phi}_t^{\mathrm{CG}}, \widehat{\mathcal{K}}\widehat{\phi}_t^{\mathrm{CG}}\rangle} \xrightarrow{\mathbb{P}} \frac{(\boldsymbol{\mu}^{\mathrm{T}}\phi_t^{\mathrm{CG}})^2}{\langle\phi_t^{\mathrm{CG}}, \mathcal{K}\phi_t^{\mathrm{CG}}\rangle} = \|\Delta_{\boldsymbol{\nu}}^{\mathrm{CG},t}\|^2 \xrightarrow{d\to\infty} \|\Delta_{\boldsymbol{\mu}}\|^2.$$

The second result follows as in the proof of Proposition 1. Under H_0 we have

$$\left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \frac{\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\phi}_t^{\mathrm{CG}}}{\langle\widehat{\phi}_t^{\mathrm{CG}}, \widehat{\mathcal{K}}\widehat{\phi}_t^{\mathrm{CG}}\rangle^{1/2}} \xrightarrow{\mathbb{P}} \left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \frac{\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\phi_t^{\mathrm{CG}}}{\langle\phi_t^{\mathrm{CG}}, \mathcal{K}\phi_t^{\mathrm{CG}}\rangle^{1/2}} \xrightarrow{\mathbb{D}} N(0,1).$$

And under H_1 we have

$$T_{\mathrm{CG}} - \left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \|\Delta_{\boldsymbol{\nu}}^{\mathrm{CG},t}\| \xrightarrow{\mathbb{P}} \left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \frac{(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\phi_t^{\mathrm{CG}}}{\langle\phi_t^{\mathrm{CG}}, \mathcal{K}\phi_t^{\mathrm{CG}}\rangle^{1/2}} \xrightarrow{\mathbb{D}} N(0,1).$$

□

Proof of Theorem 3. The asymptotic distribution under H_0 and asymptotic consistent under H_1 of the test based on $T^2_{\mathrm{PC}}(B_0)$ have been proved in the Theorem 5.3,5.4 of Horváth and Kokoszka (2012). Similar results can be established for the test based on $T^2_{\mathrm{PC}}(B_Q)$.

To show the divergence rate of $T^2_{\mathrm{PC}}(B)$, record $\phi_d^{\mathrm{PC}} = \Phi_{\mathrm{PC}}^{\mathrm{T}}\boldsymbol{\theta}_0$ and $\mathcal{K}_d = \sum_{k=1}^d \lambda_{r_k}^{-1}\psi_{r_k}\psi_{r_k}^{\mathrm{T}}$, then $\phi_d^{\mathrm{PC}} = \Phi_{\mathrm{PC}}^{\mathrm{T}}(\Phi_{\mathrm{PC}}\mathcal{K}\Phi_{\mathrm{PC}}^{\mathrm{T}})^{-1}\Phi_{\mathrm{PC}}\boldsymbol{\mu} = \mathcal{K}_d\boldsymbol{\mu}$ and $\|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2 = \boldsymbol{\mu}^{\mathrm{T}}\phi_d^{\mathrm{PC}}$. We write

$$|\widehat{\boldsymbol{\mu}}^{\mathrm{T}}\widehat{\phi}_d^{\mathrm{PC}} - \boldsymbol{\mu}^{\mathrm{T}}\phi_d^{\mathrm{PC}}| \le \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|\|\widehat{\phi}_d^{\mathrm{PC}}\| + \|\boldsymbol{\mu}\|\|\widehat{\phi}_d^{\mathrm{PC}} - \phi_d^{\mathrm{PC}}\|, \tag{S3.3}$$

$$\|\widehat{\phi}_d^{\mathrm{PC}} - \phi_d^{\mathrm{PC}}\| \le \|\widehat{\mathcal{K}}_d - \mathcal{K}_d\|_\infty\|\widehat{\boldsymbol{\mu}}\| + \|\mathcal{K}_d\|_\infty\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|, \tag{S3.4}$$

and

$$\begin{aligned}
&\|\widehat{\mathcal{K}}_d - \mathcal{K}_d\|_\infty \\
=&\|\sum_{k=1}^d \hat{\lambda}_{r_k}^{-1}\hat{\psi}_{r_k}\hat{\psi}_{r_k}^{\mathrm{T}} - \sum_{k=1}^d \lambda_{r_k}^{-1}\psi_{r_k}\psi_{r_k}^{\mathrm{T}}\|_\infty \\
\le& d\max_{1\le k\le d}\|\hat{\lambda}_{r_k}^{-1}\hat{\psi}_{r_k}\hat{\psi}_{r_k}^{\mathrm{T}} - \lambda_{r_k}^{-1}\psi_{r_k}\psi_{r_k}^{\mathrm{T}}\|_\infty \\
\le& d\max_{1\le k\le d}\{|\hat{\lambda}_{r_k}^{-1} - \lambda_{r_k}^{-1}|\|\widehat{\psi}_{r_k}\widehat{\psi}_{r_k}^{\mathrm{T}}\|_\infty + |\lambda_{r_k}^{-1}|\|\widehat{\psi}_{r_k}\widehat{\psi}_{r_k}^{\mathrm{T}} - \psi_{r_k}\psi_{r_k}^{\mathrm{T}}\|_\infty\}.
\end{aligned} \tag{S3.5}$$

By Equation (4.43) and Lemmas 4.2 and 4.3 of Bosq (2000), we have $|\hat{\lambda}_{r_k} - \lambda_{r_k}| \le \|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty$ and $\|\widehat{\psi}_{r_k} - \psi_{r_k}\| \le a_{r_k}\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty$. And since $\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty = O_p(n_0^{-1/2})$, we have $|\hat{\lambda}_{r_k}^{-1} - \lambda_{r_k}^{-1}|I\{\widehat{\lambda}_{r_k} \ge \lambda_{r_k}/2\} \le 2\lambda_{r_k}^{-2}\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty$. And since the probability of the event $\{\widehat{\lambda}_{r_k} < \lambda_{r_k}/2\}$ is bounded by $\lambda_{r_k}^{-2} O_p(n_0^{-1})$ and hence converges to 0, thus we have $|\hat{\lambda}_{r_k}^{-1} - \lambda_{r_k}^{-1}| = \lambda_{r_k}^{-2} O_p(n_0^{-1/2})$. In addition, $\|\widehat{\psi}_{r_k}\widehat{\psi}_{r_k}^{\mathrm{T}} - \psi_{r_k}\psi_{r_k}^{\mathrm{T}}\|_\infty \le 2\|\widehat{\psi}_{r_k} - \psi_{r_k}\| \le 2a_{r_k}\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty = a_{r_k} O_p(n_0^{-1/2})$.

Combining Equations (S3.3),(S3.4) and (S3.5) with the facts that $\|\widehat{\boldsymbol{\mu}}\| = O_p(1)$,$\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = O_p(n_0^{-1/2})$ and $\|\mathcal{K}_d\|_\infty = \lambda_{d^*}^{-1}$, where $d^* = \max_{1\le k\le d}\{r_k\}$, we have

$$
\begin{aligned}
&\left| \|\widehat{\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}}\|^2 - \|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2 \right| = \left| \widehat{\boldsymbol{\mu}}^{\mathrm{T}} \widehat{\phi}_d^{\mathrm{PC}} - \boldsymbol{\mu}^{\mathrm{T}} \phi_d^{\mathrm{PC}} \right| \\
&\le 2\lambda_{d^*}^{-1} O_p(n_0^{-1/2}) + d \max_{1\le k\le d}\{\lambda_{r_k}^{-2} O_p(n_0^{-1/2}) + \lambda_{r_k}^{-1} a_{r_k} O_p(n_0^{-1/2})\} \\
&= d\lambda_{d^*}^{-2} O_p(n_0^{-1/2}) + d\lambda_{d^*}^{-1} a_{d^*} O_p(n_0^{-1/2}).
\end{aligned}
$$

Thus by Assumption 4, we have $\|\widehat{\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}}\|^2 \xrightarrow{\mathbb{P}} \|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2$, and under H_1,

$$
\begin{aligned}
&T_{\mathrm{PC}}^2 - \frac{n_1 n_2}{n_1+n_2}\|\Delta_{\boldsymbol{\nu}}^{\mathrm{PC},d}\|^2 \\
=&\frac{n_1 n_2}{n_1+n_2}\sum_{k=1}^{d}\frac{((\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\widehat{\psi}_{r_k})^2}{\widehat{\lambda}_{r_k}} + \frac{n_1 n_2}{n_1+n_2}\left(\sum_{k=1}^{d}\frac{(\boldsymbol{\mu}^{\mathrm{T}}\widehat{\psi}_{r_k})^2}{\widehat{\lambda}_{r_k}} - \sum_{k=1}^{d}\frac{(\boldsymbol{\mu}^{\mathrm{T}}\psi_{r_k})^2}{\lambda_{r_k}}\right) \xrightarrow{\mathbb{D}} \chi_d^2.
\end{aligned}
$$

□

Proof of Theorem 4. Under H_0, the central limit theorem yields

$$
\left(\frac{n_1 n_2}{n_1+n_2}\right)^{\frac{1}{2}} \Phi_{\mathrm{BS}}\widehat{\boldsymbol{\mu}} \xrightarrow{\mathbb{D}} \mathbf{N}_d(\mathbf{0}, \Phi_{\mathrm{BS}}\mathcal{K}\Phi_{\mathrm{BS}}^{\mathrm{T}}),
$$

Thus we have $T^2_{\text{BS}} \xrightarrow{\mathbb{D}} \chi^2_d$. Under H_1, Combining this with the facts that $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = O_p(n_0^{-1/2})$, $\|\widehat{\mathcal{K}} - \mathcal{K}\|_\infty = O_p(n_0^{-1/2})$ and $\|\varphi_k\| = O_p(1), k = 1, \dots, d$ gives

$$\|\widehat{\Delta^{\text{BS},d}_{\boldsymbol{\nu}}}\|^2 = \frac{n_1 + n_2}{n_1 n_2} T^2_{\text{BS}} \xrightarrow{\mathbb{P}} (\Phi_{\text{BS}}\boldsymbol{\mu})^{\text{T}} (\Phi_{\text{BS}} \mathcal{K} \Phi^{\text{T}}_{\text{BS}})^{-1} \Phi_{\text{BS}} \boldsymbol{\mu} = \|\Delta^{\text{BS},d}_{\boldsymbol{\nu}}\|^2.$$

And if $\varphi_k (k = 1, 2, \dots)$ generate the range of $\mathcal{K}$, then by Proposition 2.1, $\|\Delta^{\text{BS},d}_{\boldsymbol{\nu}}\|^2 \overset{d\to\infty}{\to} \|\Delta_{\boldsymbol{\mu}}\|^2$. □

Proof of Proposition 2. The conjugate gradient method minimizes the quadratic objective function in the Krylov subspace $K_d(\mathcal{K}, \boldsymbol{\mu})$ whose elements are in the form $\phi = \sum_{k=0}^{d-1} \beta_k \mathcal{K}^k \boldsymbol{\mu} = p(\mathcal{K})\boldsymbol{\mu}$, where $p(\cdot)$ is a polynomial of order lower than d. Then $\phi \in K_d(\mathcal{K}, \boldsymbol{\mu})$ can be written as $\phi = \sum_{k=0}^{d-1} \beta_k (\sum_{r=1}^{\infty} \lambda_r \psi_r \psi_r^{\text{T}})^k \boldsymbol{\mu} = \sum_{r=1}^{\infty} \sum_{k=0}^{d-1} \beta_k \lambda_r^k \psi_r \psi_r^{\text{T}} \boldsymbol{\mu} = \sum_{k=1}^{\infty} p(\lambda_k) b_k \psi_k$ with $b_k = \boldsymbol{\mu}^{\text{T}} \psi_k$. The information extracted by ϕ equals

$$\|\Delta^{\phi}_{\boldsymbol{\nu}}\|^2 = \frac{(\boldsymbol{\mu}^{\text{T}}\phi)^2}{\phi^{\text{T}} \mathcal{K} \phi} = \frac{(\sum_{k=1}^{\infty} b_k^2 p(\lambda_k))^2}{\sum_{k=1}^{\infty} b_k^2 \lambda_k p^2(\lambda_k)} = \frac{(\sum_{k=1}^{\infty} \frac{b_k^2}{\lambda_k} q(\lambda_k))^2}{\sum_{k=1}^{\infty} \frac{b_k^2}{\lambda_k} q^2(\lambda_k)} < \sum_{k=1}^{\infty} \frac{b_k^2}{\lambda_k},$$

where $q(\lambda) = p(\lambda)\lambda$ is a polynomial of degree at most d such that $q(0) = 0$. The last inequality is according to the Jensen's inequality, where the equality is achieved if and only if $q(\lambda_1) = q(\lambda_2) = \cdots$, which is impossible because the degree of $p(\lambda)$ is less than d.

Record $c_1 = \max(\arg\max_c |\sum_k I(q(\lambda_k) = c)|)$, and define $A_1 = \{k \mid q(\lambda_k) = c_1\}$, $A_2 = \{k \mid q(\lambda_k) \neq c_1\}$, where $|A_1| \leq d-1$. Recored $a_k = b_k^2/\lambda_k, q_k = q(\lambda_k)$, $\sum_{k\in A_i} a_k = e_i$ and $\sum_{k\in A_2} a_k q_k / e_2 = c_2$, then $\sum_{k\in A_i} a_k q_k = c_i e_i$ for $i = 1, 2$, $\sum_{k\in A_1} a_k q_k^2 = c_1^2 e_1$, $\sum_{k\in A_2} a_k q_k^2 > c_2^2 e_2$, and

$$\|\Delta_{\boldsymbol{\nu}}^{\phi}\|^2 = \frac{(\sum_{k=1}^{\infty} a_k q_k)^2}{\sum_{k=1}^{\infty} a_k q_k^2} < \frac{(c_1 e_1 + c_2 e_2)^2}{c_1^2 e_1 + c_2^2 e_2}.$$

As c_2 is the weighted average of q_k, where $k \in A_2$, $q_k \to 0$ as $\lambda_k \to 0$ by the continuity of $q(\cdot)$, the weight $a_k < e_1 + e_2 = \|\Delta_{\boldsymbol{\mu}}\|^2 < \infty$ and a_k doesn't decline to 0 as $k \to \infty$, we have $c_2 \to 0$ and $\|\Delta_{\boldsymbol{\nu}}^{\phi}\|^2 < e_1$ when there are infinite eigenvalues close to 0. Thus the information extracted by ϕ is less than $\sum_{k\in A_1} b_k^2/\lambda_k$.

The information extracted by $\Phi_{\text{PC}}(B_Q)$ is $\sum_{k=1}^{d} b_{r_k}^2/\lambda_{r_k}$, which is larger than $\sum_{k\in A_1} b_k^2/\lambda_k$ due to the way we choose Q. Thus for a given d, the testing based on $\Phi_{\text{PC}}(B_Q)$ is more powerful than based on ϕ_d^{CG}. In addition, by the proof of Proposition 2 in Kraus and Stefanucci (2019), we have the testing based on ϕ_d^{CG} is more powerful than based on $\Phi_{\text{PC}}(B_0)$. □

References

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, Volume 149. Springer Science & Business Media.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Kraus, D. and M. Stefanucci (2019). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika 106*(1), 161–180.