Supplementary Material for "Linear Model Selection when Covariates Contain Errors"

Xinyu Zhang
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, xinyu@amss.ac.cn

Haiying Wang
Department of Mathematics and Statistics, University of New Hampshire, Durham, NH 03824, haiying.wang@unh.edu

Yanyuan Ma
Department of Statistics, Penn State University, State College, PA 16802, yanyuanma@gmail.com

Raymond J. Carroll
Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, and School of Mathematical and Physical Sciences, University of Technology Sydney, Broadway NSW 2007, carroll@stat.tamu.edu

S.1 Additional simulation using the covariance structure and covariates of the empirical data

Following the request of a referee, we perform additional simulation where the covariance structure and covariate distribution resemble those of the empirical data. Specifically, we generate four continuous covariates, $X_{i,1}, \dots, X_{i,4}$, from Normal$(\mathbf{0}, \sigma_x^2 \boldsymbol{\Omega})$, where σ_x is chosen the same as in

Example 1 and

$$\mathbf{\Omega} = \begin{pmatrix} 1 & 0.658 & 0.579 & 0.065 \\ 0.658 & 1 & 0.607 & -0.057 \\ 0.579 & 0.607 & 1 & -0.071 \\ 0.065 & -0.057 & -0.071 & 1 \end{pmatrix},$$

the correlation structure from the four continuous covariates in the empirical data, and two discrete covariates $X_{i,5}$ and $X_{i,6}$ through re-sampling from the empirical data. We add measurement errors the first three covariates. Note that as in the data, the last three covariates do not contain measurement error, so $W_{i,4} = X_{i,4}, W_{i,5} = X_{i,5}$ and $W_{i,6} = X_{i,6}$. All other settings are the same as those with Example 1 and Example 2 respectively. Tables S.1 and S.2 contain the simulation results, which demonstrate the superior performance of UBERIC.

S.2 Model selection based on estimation

Our model selection method is prediction-based in the sense that it selects a model to minimize the loss in estimating $\boldsymbol{\mu}$, the mean of $\mathbf{Y}$. We now compare its performance with an estimation-based model selection procedure.

The focused information criterion (FIC) proposed by Claeskens and Hjort (2003) selects a model based on estimation. Specifically, the framework used to develop FIC is

$$Y_i = \mathbf{X}_{1,i}^{\mathrm{T}}\boldsymbol{\beta}_1 + \mathbf{X}_{2,i}^{\mathrm{T}}\boldsymbol{\beta}_2 + \epsilon_i, \qquad \boldsymbol{\beta}_2 = \boldsymbol{\delta}/\sqrt{n}, \tag{S.1}$$

where $\boldsymbol{\beta}_1$ is a vector containing parameters corresponding to the "needed" covariates, $\boldsymbol{\beta}_2$ is a vector of parameters corresponding to the "unneeded" predictors, and $\boldsymbol{\beta}_2 = \boldsymbol{\delta}/\sqrt{n}$ is a local misspecification assumption. Some or all components of $\mathbf{X}_i = (\mathbf{X}_{1,i}^{\mathrm{T}}, \mathbf{X}_{2,i}^{\mathrm{T}})^{\mathrm{T}}$ are measured with error, i.e., $\mathbf{W}_i = \mathbf{X}_i + \mathbf{U}_i$, and $\mathbf{U}_i$ is a mean zero normal random vector with variance-covariance matrix $\mathbf{\Sigma}$. Assume that the parameter of interest is $\mu = \mu(\boldsymbol{\beta})$. Let $\mathbf{\Pi}_{(s)}^{\mathrm{T}}$ be a selection matrix with the element matching $\boldsymbol{\beta}_{2,(s)}$ so that $\boldsymbol{\beta}_{2,(s)} = \mathbf{\Pi}_{(s)}^{\mathrm{T}}\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_{(s)} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \mathbf{\Pi}_{(s)}^{\mathrm{T}}\boldsymbol{\beta}_2^{\mathrm{T}})^{\mathrm{T}}$. Denote $\mathbf{B}_n = \sum_{i=1}^n \mathbf{W}_i\mathbf{W}_i^{\mathrm{T}} - n\widehat{\mathbf{\Sigma}} = \begin{pmatrix} \mathbf{B}_{n11} & \mathbf{B}_{n12} \\ \mathbf{B}_{n21} & \mathbf{B}_{n22} \end{pmatrix}$. Using an approach similar to that in Wang et al. (2016), we obtain the FIC for estimating $\mu = \mu(\boldsymbol{\beta})$ as

$$\mathrm{FIC}_{n(s)} = \left\{\widehat{\omega}^{\mathrm{T}}(\mathbf{I} - \mathbf{A}_n^{-1/2}\mathbf{H}_{n(s)}\mathbf{A}_n^{1/2})\widehat{\boldsymbol{\delta}}\right\}^2 - 2(\mathbf{0},\ \widehat{\omega}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{H}_{n(s)}\mathbf{A}_n^{1/2})\widehat{\mathbf{P}}_n\widehat{\boldsymbol{\mu}}_{\boldsymbol{\beta}},$$

where $\widehat{\omega} = (\mathbf{B}_{n21}\mathbf{B}_{n11}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}_1},\ -\mathbf{I})\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \partial\mu/\partial\boldsymbol{\beta}$, $\mathbf{H}_{n(s)} = \mathbf{A}_n^{1/2}\boldsymbol{\Pi}_{(s)}\left(\boldsymbol{\Pi}_{(s)}^{\mathrm{T}}\mathbf{A}_n\boldsymbol{\Pi}_{(s)}\right)^{-1}\boldsymbol{\Pi}_{(s)}^{\mathrm{T}}\mathbf{A}_n^{1/2}$, $\mathbf{A}_n = \mathbf{B}_{n22} - \mathbf{B}_{n21}\mathbf{B}_{n11}^{-1}\mathbf{B}_{n12}$, $\widehat{\boldsymbol{\delta}} = \sqrt{n}\widehat{\boldsymbol{\beta}}_2 = \mathbf{A}_n^{-1}\sum_{i=1}^n\left(\mathbf{W}_{2,i} - \mathbf{B}_{n21}\mathbf{B}_{n11}^{-1}\mathbf{W}_{1,i}\right)Y_i$, $\widehat{\mathbf{P}} = n\mathbf{B}_n^{-1}\widehat{\mathbf{F}}\mathbf{B}_n^{-1}$, and

$$\widehat{\mathbf{F}} = \sum_{i=1}^n \left[\mathbf{W}_i\left\{Y_i - \mathbf{W}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right\} + \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\beta}}\right]^{\otimes 2},$$

where $\mathbf{R}^{\otimes 2} = \mathbf{R}\mathbf{R}^{\mathrm{T}}$ for any matrix $\mathbf{R}$. The FIC procedure selects the candidate model with the smallest value of $\mathrm{FIC}_{n(s)}$.

We implemented FIC in Example I of the main paper, with results in Table S.3. It is seen that when n is small (25 or 50), UBERIC with $\lambda_n = 2$ performs the best, and when n is large (100 or 200), UBERIC with $\lambda_n = \log(n)pn^{1/2}$ has the best performance. Regardless the sample size, UBERIC outperforms FIC in most cases in terms of relative loss.

We also performed a simulation following model (S.1) to compare the performance of UBERIC and FIC in terms of estimation. We set $\boldsymbol{\beta}_1 = 2$, $\boldsymbol{\beta}_2 = (1.5, 2, 0, -1)^{\mathrm{T}}/\sqrt{n}$ and compare MSE of the estimation of $\boldsymbol{\beta}_1$ using 500 replications. For easy comparison, we scale all MSEs so that the MSE for FIC is 1. From the results in Table S.4, we see that UBERIC with $\lambda_n = 2$ performs the best in most cases. When $n = 25$ and σ is small, UBERIC with $\lambda_n = \log(n)pn^{1/2}$ is the winner but when $n = 100$ and 200, FIC outperforms UBERIC with $\lambda_n = \log(n)pn^{1/2}$. It is encouraging that the prediction-based UBERIC method is competitive with FIC in terms of estimation as well.

Lastly, we implemented FIC in the data example as well. The boxplots of the squared prediction errors are in Figure S.1. In this example, UWERIC outperforms FIC slightly.

S.3 Combine UBERIC and the penalized least squares variable selection procedure

When p is small, the optimization involved in UBERIC can be performed in a brute-force way. However, to facilitate the computation when p is large, we propose to combine UBERIC and the penalized least squares (PLS) variable selection procedure proposed by Liang and Li (2009). We use the PLS with multiple turning parameter values to select multiple different models and use the selected models as candidate models to feed to UBERIC. Following Liang and Li (2009), we use 50 turning parameter values in a grid, so the number of candidate models selected by the PLS is not larger than 50, which, for the case of a large p, is much smaller than 2^p, the number of all possible models, and thus this will significantly improve the computation efficiency.

To illustrate the performance of the combined method, we carried out experiments and added the results in Tables S.5 and S.6. These results suggest that UBERIC$_{PLS}$ sometimes performs even better than UBERIC.

S.4 Proof of Lemma 2

The result in (7) is a generalization of Lemma 1, and it can be obtained by the following calculation.

$$\begin{aligned} E\text{tr}(\mathbf{M}_1\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}) &= E\{\text{vec}(\mathbf{M}_1^{\mathrm{T}})^{\mathrm{T}}\text{vec}(\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\} \\ &= E\sum_{i=1}^{np_{(s)}}\text{vec}(\mathbf{M}_1^{\mathrm{T}})_i\text{vec}(\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})_i = E\sum_{i=1}^{np_{(s)}}\frac{\partial\text{vec}(\mathbf{M}_1^{\mathrm{T}})_i}{\partial\text{vec}(\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})_i} = E\text{tr}(\mathrm{D}\mathbf{M}_1^{\mathrm{T}}), \end{aligned}$$

where the third equality is from Lemma 1.

The result in (8) is a direct result from (xiii) of Theorem 3.1 in Magnus and Neudecker (1979).

For (9),

$$\begin{aligned} &\text{tr}\left\{\frac{\mathbf{M}_4\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{M}_3}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} = \text{tr}\left\{\frac{\text{vec}(\mathbf{M}_4\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{M}_3)}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} = \text{tr}\left\{\frac{(\mathbf{M}_3^{\mathrm{T}}\otimes\mathbf{M}_4)\text{vec}(\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} \\ &= \text{tr}\left\{(\mathbf{M}_3^{\mathrm{T}}\otimes\mathbf{M}_4)\right\} = \text{tr}(\mathbf{M}_3)\text{tr}(\mathbf{M}_4). \end{aligned}$$

The result in (10) is proved as

$$\begin{aligned} &\text{tr}\left\{\frac{\mathbf{M}_1^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{M}_2^{\mathrm{T}}}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} = \text{tr}\left\{\frac{(\mathbf{M}_2\otimes\mathbf{M}_1^{\mathrm{T}})\text{vec}(\mathrm{d}\breve{\mathbf{U}}_{(s)})}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} \\ &= \text{tr}\left\{\frac{(\mathbf{M}_2\otimes\mathbf{M}_1^{\mathrm{T}})K_{p_{(s)}n}\text{vec}(\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} = \text{tr}\left\{(\mathbf{M}_2\otimes\mathbf{M}_1^{\mathrm{T}})K_{p_{(s)}n}\right\} = \text{tr}(\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_2), \end{aligned}$$

where the second equality is from a property of commutation matrices (Magnus and Neudecker, 1979).

For (11),

$$\text{tr}\left\{\frac{\text{tr}(\mathbf{M}_1\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{M}_3)\mathbf{M}_2^{\mathrm{T}}}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\} = \text{tr}\left\{\frac{\text{vec}(\mathbf{M}_2^{\mathrm{T}})\text{tr}(\mathbf{M}_3\mathbf{M}_1\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}}\right\}$$

$$
= \text{tr}\left\{ \frac{\text{vec}(\mathbf{M}_2^{\mathrm{T}})\text{vec}(\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_3^{\mathrm{T}})^{\mathrm{T}}\text{vec}(\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})}{\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}} \right\}
$$

$$
= \text{tr}\left\{\text{vec}(\mathbf{M}_2^{\mathrm{T}})\text{vec}(\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_3^{\mathrm{T}})^{\mathrm{T}}\right\} = \text{tr}\{\text{vec}(\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_3^{\mathrm{T}})^{\mathrm{T}}\text{vec}(\mathbf{M}_2^{\mathrm{T}})\} = \text{tr}\left(\mathbf{M}_3\mathbf{M}_1\mathbf{M}_2^{\mathrm{T}}\right).
$$

□

S.5 Proof of Theorem 2

Let $\widehat{\boldsymbol{\epsilon}}_{(s)} = (\widehat{\epsilon}_{(s),1}, \ldots, \widehat{\epsilon}_{(s),n})^{\mathrm{T}} = \mathbf{Y} - \mathbf{H}_{(s)}\mathbf{Y}$. Notice that

$$
\begin{aligned}
E\|\widehat{\boldsymbol{\mu}}_{(s)} - \mathbf{Y}\|^2 &= E\|\mathbf{X}\mathbf{G}_{(s)}\mathbf{Y} - \mathbf{Y}\|^2 = E\|\widehat{\boldsymbol{\epsilon}}_{(s)} + \mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\|^2 \\
&= E\|\widehat{\boldsymbol{\epsilon}}_{(s)}\|^2 + E\|\mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\|^2 + 2E(\widehat{\boldsymbol{\epsilon}}_{(s)}^{\mathrm{T}}\mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}). \qquad \text{(S.2)}
\end{aligned}
$$

The last two terms in (S.2) contains unobserved $\mathbf{U}_{(s)}$. We deal with them separately in the following. For the second term in (S.2),

$$
\begin{aligned}
&E\|\mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\|^2 = E\|\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\|^2 = E(\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}) \\
&= E\text{tr}(\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}) = E\text{tr}\{\mathrm{D}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\}. \qquad \text{(S.3)}
\end{aligned}
$$

For the matrix product in (S.3), we first find its differential,

$$
\begin{aligned}
&\mathrm{d}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}) \\
=&\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&+ \boldsymbol{\Sigma}_{(s)}^{1/2}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&+ \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)} - \mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)} - \mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
=&\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&+ \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&- \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&- \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&+ \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&- \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}
\end{aligned}
$$

$$- \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}. \tag{S.4}$$

Then from (9) and (10) in Lemma 2, and equations (12) and (S.4),

$$\begin{aligned}
&E\mathrm{tr}\{\mathrm{D}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\}\\
=&nE(\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y})\\
&+E\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\\
&-E\mathrm{tr}(\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\\
&-E\mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\\
&+E\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\\
&-E\mathrm{tr}(\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\\
&-E\mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\\
&\equiv \Delta_1+\Delta_2-\Delta_3-\Delta_4+\Delta_5-\Delta_6-\Delta_7.
\end{aligned} \tag{S.5}$$

In (S.5), Δ_1 does not involve $\breve{\mathbf{U}}_{(s)}$, so we just need to deal with the other 6 terms. For the second term,

$$\Delta_2 = E\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}) = E\mathrm{tr}[\mathrm{D}\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})\}]. \tag{S.6}$$

The differential of the matrix product in (S.6) is

$$\begin{aligned}
&\mathrm{d}\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}\\
=&\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathrm{d}\boldsymbol{\Lambda}_{(s)})(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})\\
=&-\mathrm{tr}\{\boldsymbol{\Sigma}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}+\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)})\}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})\\
&+\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\{\boldsymbol{\Sigma}_{(s)}^{1/2}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}-\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}-\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}\\
=&-2\mathrm{tr}\{\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})\\
&+\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\{\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}\\
&-\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\{\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}\\
&-\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\{\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}.
\end{aligned} \tag{S.7}$$

So from (9), (10) and (11) of Lemma 2, and equations (S.6) and (S.7), we have

$$E\Delta_2 = \{-2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}+\mathrm{tr}^2(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{Y}$$

$$- \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\}. \tag{S.8}$$

Similarly, we obtain results for terms Δ_3, ..., Δ_7 in (S.5) as follows. For the third term in (S.5),

$$\begin{aligned}
\Delta_3 &= E\mathrm{tr}[\mathrm{D}\{\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\}] \\
&= E\mathrm{tr}[\mathrm{D}\{\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)})\}] \\
&= E\mathrm{tr}[\{\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathrm{d}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{W}_{(s)}\mathbf{G}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{W}_{(s)}\mathrm{d}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
&= E\mathrm{tr}[\{\mathrm{tr}\{\mathbf{\Sigma}_{(s)}(-\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{\Lambda}_{(s)})\}(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\{\mathbf{\Sigma}_{(s)}^{1/2}(\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&\qquad\qquad - \mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\} \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\{\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\} \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\{\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{W}_{(s)}(\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&\qquad\qquad - \mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
&= E\mathrm{tr}[\{-2\mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\}(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad - \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad - \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}) \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}) \\
&\qquad - \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&\qquad - \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})(\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\mathbf{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
&= E[-2\mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}) + \mathrm{tr}^2(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} \\
&\qquad - \mathrm{tr}^2(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} - 2\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{H}_{(s)}\mathbf{Y} \\
&\qquad + \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} + n\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} \\
&\qquad - \mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})\mathrm{tr}(\mathbf{H}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}].
\end{aligned} \tag{S.9}$$

For the forth term in (S.5),

$$
\begin{aligned}
\Delta_4 =& E\mathrm{tr}\{\mathrm{D}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\} \\
=& E\mathrm{tr}\mathrm{D}(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
=& E\mathrm{tr}[\{(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&+ (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&+ (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathrm{d}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E\mathrm{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\qquad - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\} \\
&+ \{\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)} - \mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)} \\
&\qquad - \mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\} \\
&+ \{\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&\qquad - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E\mathrm{tr}[\{(\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&+ (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}) \\
&+ (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}) \\
&- (\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} - \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{H}_{(s)}\mathbf{Y} \\
&- 2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} + \mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} \\
&- \mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} - \mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} \\
&+ n\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} - \mathrm{tr}(\mathbf{H}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\}. \qquad \text{(S.10)}
\end{aligned}
$$

For the fifth term in (S.5),

$$
\begin{aligned}
\Delta_5 =& E\text{tr}\{\mathrm{D}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\} \\
=& E\text{tr}\{(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}})/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\} \\
=& E\text{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}(-\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)})\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\qquad\qquad - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E\text{tr}[\{-\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}} \\
&\quad - \boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\} \\
&\quad - \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\} \\
&\quad - \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E\{-\text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} + \text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{Y} \\
&\quad - \text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} - 2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\}. \qquad \text{(S.11)}
\end{aligned}
$$

For the sixth term in (S.5),

$$
\begin{aligned}
\Delta_6 =& E\text{tr}\{\mathrm{D}(\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\} \\
=& E\text{tr}\{(\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)} + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{W}_{(s)}\mathbf{G}_{(s)} + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{W}_{(s)}\mathrm{d}\mathbf{G}_{(s)})/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\} \\
=& E\text{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}(-\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)})\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\qquad\qquad - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)} \\
&\quad + \boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{W}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} \\
&\qquad\qquad - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}] \\
=& E[-\text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} + \text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} \\
&\quad - \text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} + \mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}
\end{aligned}
$$

$$
\begin{aligned}
&+ n\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} - \mathrm{tr}(\mathbf{H}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\\
&- 3\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}]. \qquad \text{(S.12)}
\end{aligned}
$$

For the seventh term in (S.5),

$$
\begin{aligned}
\Delta_7 =& E\mathrm{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\\
=& E\mathrm{tr}[\mathrm{D}\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\}]\\
=& E\mathrm{tr}[\mathrm{d}\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2})^{\mathrm{T}}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=& E\mathrm{tr}[\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)} + \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&\qquad + \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\mathbf{G}_{(s)}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=& E\mathrm{tr}([\mathrm{tr}\{\boldsymbol{\Sigma}_{(s)}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}} - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\\
&\qquad - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&+ \mathrm{tr}\{\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)} - \mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\\
&\qquad - \mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&+ \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\\
&\qquad - \boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)} - \mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})]/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}})\\
=& E\mathrm{tr}[\{\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&+ \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&+ \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=& E\mathrm{tr}[2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} - 2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{H}_{(s)}\mathbf{Y}\\
&- 2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} + n\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\\
&- \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathrm{tr}(\mathbf{H}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}
\end{aligned}
$$

$$- \text{tr}(\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}]. \tag{S.13}$$

Now we consider the last term in (S.2). Using an approach similar to that used to deal with Δ_1 in (S.5), we have

$$\begin{aligned}
&2E(\widehat{\boldsymbol{\epsilon}}_{(s)}^{\mathrm{T}}\mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y})\\
=&2E\{(\mathbf{Y}-\mathbf{H}_{(s)}\mathbf{Y})^{\mathrm{T}}\mathbf{U}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\}\\
=&2E\{\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}(\mathbf{Y}-\mathbf{H}_{(s)}\mathbf{Y})\}\\
=&2E\text{tr}\{(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\}\\
=&2E\text{tr}[\mathrm{D}\{(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\}^{\mathrm{T}}]\\
=&2E\text{tr}[\mathrm{D}\{\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\}]\\
=&2E\text{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\mathrm{d}\mathbf{W}_{(s)}^{\mathrm{T}}-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}^{\mathrm{T}}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=&2E\text{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}(\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}-\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\\
&\qquad\qquad\qquad-\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)})\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}-\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\\
&\qquad\qquad\qquad-\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}})\mathbf{W}_{(s)}^{\mathrm{T}}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=&2E\text{tr}[\{\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\\
&\qquad-\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&\qquad+\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathrm{d}\breve{\mathbf{U}}_{(s)}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\\
&\qquad+\boldsymbol{\Sigma}_{(s)}^{1/2}\mathbf{G}_{(s)}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}^{1/2}\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}\mathbf{H}_{(s)}\}/\mathrm{d}\breve{\mathbf{U}}_{(s)}^{\mathrm{T}}]\\
=&2E\{\text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{Y}-\text{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{H}_{(s)}\mathbf{Y}\\
&\qquad-\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n-\mathbf{H}_{(s)})\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}-n\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\\
&\qquad-\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}+\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}
\end{aligned}$$

$$
\begin{aligned}
&\quad + \mathrm{tr}(\mathbf{H}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}\} \\
=&2E[\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - 2\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} + \mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} \\
&\quad + 2\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{H}_{(s)}\mathbf{Y} + \{\mathrm{tr}(\mathbf{H}_{(s)}) - (n+2)\}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y}]. \qquad \text{(S.14)}
\end{aligned}
$$

Combining (S.2), (S.3), (S.5), (S.8), (S.9), (S.10), (S.11), (S.12), (S.13) and (S.14), we have

$$
\begin{aligned}
&E\|\widehat{\boldsymbol{\mu}}_{(s)} - \mathbf{Y}\|^2 \\
=&E\Bigg(\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{H}_{(s)})\mathbf{Y} \\
&\quad + \mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{H}_{(s)})^2\mathbf{Y}[\mathrm{tr}\{(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})^2\} + \mathrm{tr}^2(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}) + 2\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})] \\
&\quad + 2\{\mathrm{tr}(\mathbf{H}_{(s)}) - n - 4\}\mathbf{Y}^{\mathrm{T}}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} \\
&\quad + 8\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)}\mathbf{Y} \\
&\quad + \frac{4}{n^2}(\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} - 2\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^3\mathbf{Y} + \mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^4\mathbf{Y}) \\
&\quad - \frac{4\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}) + 4}{n}(\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^3\mathbf{Y}) \\
&\quad - (\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y})\Bigg[\frac{2\{\mathrm{tr}(\mathbf{H}_{(s)}) - n - 2\}\mathrm{tr}(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}) - 4}{n} \\
&\qquad\qquad + \frac{2\mathrm{tr}(\mathbf{H}_{(s)}^2) + 2(n-1)\mathrm{tr}(\mathbf{H}_{(s)})}{n^2}\Bigg]\Bigg). \qquad \text{(S.15)}
\end{aligned}
$$

It follows from (15) that

$$
\mathrm{tr}\{(\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)})^2\} = \frac{1}{n^2}\mathrm{tr}\{(\mathbf{W}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)} - \mathbf{I}_{p_{(s)}})^2\} = \frac{t_{(s),2} - 2t_{(s),1} + p_{(s)}}{n^2} \qquad \text{(S.16)}
$$

and

$$
\begin{aligned}
&\mathbf{G}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\mathbf{G}_{(s)} \\
&= \mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)}^{\mathrm{T}}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\boldsymbol{\Sigma}_{(s)}\boldsymbol{\Lambda}_{(s)}\mathbf{W}_{(s)} \\
&= \frac{1}{n^2}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)}(\mathbf{W}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)} - \mathbf{I}_{p_{(s)}})^2\mathbf{W}_{(s)} \\
&= \frac{1}{n^2}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)}(\mathbf{W}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)}\mathbf{W}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)} - 2\mathbf{W}_{(s)}^{\mathrm{T}}\mathbf{W}_{(s)}\boldsymbol{\Lambda}_{(s)} + \mathbf{I}_{p_{(s)}})\mathbf{W}_{(s)} \\
&= \frac{1}{n^2}(\mathbf{H}_{(s)} - \mathbf{I}_n)^2\mathbf{H}_{(s)}. \qquad \text{(S.17)}
\end{aligned}
$$

Replacing $\mathrm{tr}(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})$, $\mathrm{tr}\{(\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)})^2\}$ and $\mathbf{G}_{(s)}^{\mathrm{T}}\mathbf{\Sigma}_{(s)}\mathbf{\Lambda}_{(s)}\mathbf{\Sigma}_{(s)}\mathbf{G}_{(s)}$ in (S.15) by (17), (S.16) and (S.17), respectively, and after some algebra manipulations, $E\|\widehat{\boldsymbol{\mu}}_{(s)} - \mathbf{Y}\|^2$ is equal to

$$
\begin{aligned}
&E\Big[\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{H}_{(s)})\mathbf{Y} \\
&\qquad + n^{-2}\{a_{(s)}^2 + (2n-1)a_{(s)} + t_2 - t_1\}\mathbf{Y}^{\mathrm{T}}\mathbf{Y} \\
&\qquad + n^{-2}\{-2a_{(s)}^2 + a_{(s)}(-2t_1 - 2n + 6) - 4t_2 + t_1(-2n+6) + 2n - 8\}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}\mathbf{Y} \\
&\qquad + n^{-2}\{a_{(s)}^2 + a_{(s)}(2t_1 - 9) + t_1(2n-7) + 3t_2 - 4n + 28\}\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^2\mathbf{Y} \\
&\qquad + n^{-2}(4a_{(s)} + 2t_1 + 2n - 32)\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^3\mathbf{Y} \\
&\qquad + n^{-2}12\mathbf{Y}^{\mathrm{T}}\mathbf{H}_{(s)}^4\mathbf{Y}\Big] \\
&= E\{\mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{H}_{(s)})\mathbf{Y} + B_{(s)}\},
\end{aligned}
$$

which completes the proof.

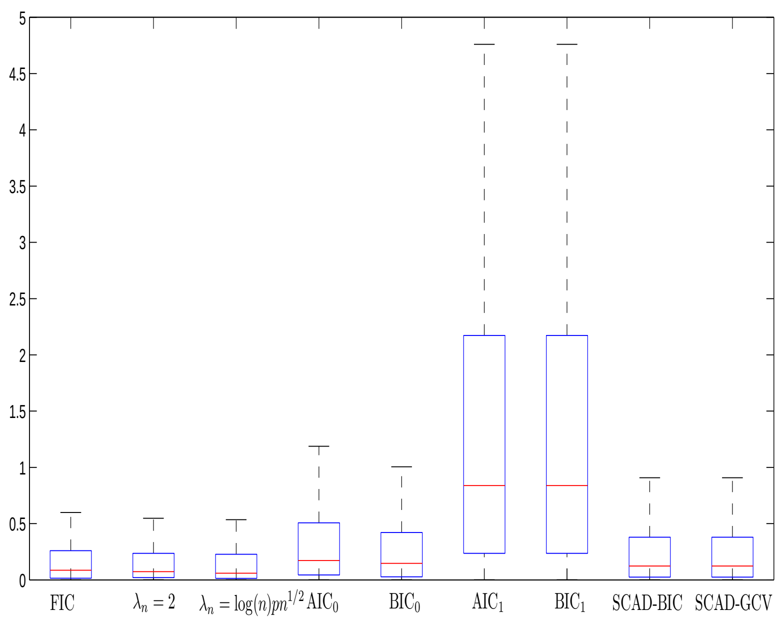

Figure S.1: Analysis of WISH data. Boxplots of 192 squared prediction errors. Methods compared are FIC, UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, AIC_0, BIC_0, AIC_1, BIC_1, SCAD-BIC, and SCAD-GCV.

Table S.1: Simulation results of Example I *using the covariance structure and covariates from the empirical data*: relative loss. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, AIC_0, BIC_0, AIC_1, BIC_1, SCAD-BIC, and SCAD-GCV. The best results are in bold face.

n	σ	τ	$\lambda_n = 2$	$\lambda_n = \log(n)pn^{1/2}$	AIC_0	BIC_0	AIC_1	BIC_1	SCAD-BIC	SCAD-GCV
25	0.5	0.85	2.044	5.285	1.897	**1.740**	6.502	6.162	7.584	7.371
		0.95	1.646	3.150	1.860	**1.627**	2.067	1.984	2.086	1.973
	1	0.85	2.239	5.907	2.111	**1.941**	5.362	5.109	7.642	7.653
		0.95	1.895	4.678	2.021	**1.774**	2.335	2.230	2.185	1.928
50	0.5	0.85	**1.412**	4.246	2.007	1.874	1.996	1.941	3.448	3.271
		0.95	**1.383**	1.472	1.673	1.470	1.631	1.567	1.795	1.884
	1	0.85	**1.536**	5.725	1.933	1.762	2.019	1.930	3.480	2.967
		0.95	1.454	2.442	1.619	**1.410**	1.739	1.657	3.106	2.029
100	0.5	0.85	**1.233**	2.612	2.069	1.987	1.394	1.371	2.860	2.263
		0.95	1.219	**1.030**	1.610	1.468	1.336	1.305	3.402	2.897
	1	0.85	**1.328**	5.282	2.025	1.915	1.515	1.464	2.750	2.373
		0.95	1.292	1.100	1.613	1.425	1.416	1.363	3.992	3.402
200	0.5	0.85	**1.148**	1.243	2.142	2.106	1.213	1.202	3.307	2.792
		0.95	1.134	**1.032**	1.576	1.488	1.185	1.173	4.942	3.205
	1	0.85	**1.198**	3.681	2.077	2.012	1.273	1.241	2.706	2.473
		0.95	1.175	**1.027**	1.571	1.456	1.232	1.202	6.442	3.348

Table S.2: Simulation results of Example II *using the covariance structure and covariates from the empirical data*: frequency in selecting the smallest correct model. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, AIC_0, BIC_0, AIC_1, BIC_1, SCAD-BIC, and SCAD-GCV. The best results are in bold face.

n	σ	τ	$\lambda_n = 2$	$\lambda_n =$ $\log(n)pn^{1/2}$	AIC_0	BIC_0	AIC_1	BIC_1	SCAD-BIC	SCAD-GCV
25	0.5	0.85	0.516	0.320	0.388	**0.586**	0.098	0.172	0.440	0.436
		0.95	0.524	**0.882**	0.470	0.702	0.110	0.196	0.628	0.602
	1	0.85	0.506	0.278	0.404	**0.618**	0.118	0.202	0.400	0.398
		0.95	0.550	**0.808**	0.476	0.708	0.138	0.234	0.612	0.596
50	0.5	0.85	0.522	**0.644**	0.210	0.530	0.056	0.158	0.626	0.618
		0.95	0.510	**0.964**	0.378	0.696	0.074	0.192	0.662	0.630
	1	0.85	0.506	0.452	0.278	0.552	0.094	0.204	0.604	**0.612**
		0.95	0.504	**0.966**	0.382	0.730	0.108	0.222	0.680	0.626
100	0.5	0.85	0.486	**0.892**	0.110	0.432	0.040	0.154	0.802	0.780
		0.95	0.484	**0.986**	0.322	0.714	0.044	0.146	0.808	0.762
	1	0.85	0.460	0.676	0.168	0.500	0.064	0.178	**0.798**	0.782
		0.95	0.444	**0.992**	0.326	0.740	0.070	0.210	0.806	0.756
200	0.5	0.85	0.502	**0.990**	0.022	0.162	0.022	0.116	0.880	0.856
		0.95	0.506	**0.994**	0.178	0.618	0.016	0.120	0.916	0.846
	1	0.85	0.512	0.876	0.040	0.272	0.038	0.244	**0.892**	0.862
		0.95	0.500	**0.998**	0.220	0.670	0.050	0.262	0.904	0.844

Table S.3: Simulation results under the setup in Example I of the main paper: relative loss. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, and FIC. The best results are in bold face.

n	σ	τ	$\lambda_n = 2$	$\lambda_n = \log(n)pn^{1/2}$	FIC
25	0.5	0.85	**1.390**	4.094	3.611
		0.95	**1.412**	1.450	1.693
	1	0.85	**1.511**	5.321	3.315
		0.95	**1.465**	1.774	1.794
50	0.5	0.85	**1.350**	2.899	1.928
		0.95	1.356	**1.069**	1.485
	1	0.85	**1.496**	4.971	2.164
		0.95	1.446	**1.068**	1.629
100	0.5	0.85	**1.261**	1.512	1.413
		0.95	1.192	**1.015**	1.254
	1	0.85	**1.348**	3.417	1.540
		0.95	1.235	**1.008**	1.316
200	0.5	0.85	1.161	**1.006**	1.205
		0.95	1.111	**1.006**	1.130
	1	0.85	**1.205**	1.553	1.261
		0.95	1.139	**1.003**	1.163

Table S.4: Simulation results for $\boldsymbol{\beta}_1 = 2$ and $\boldsymbol{\beta}_2 = (1.5, 2, 0, -1)^{\mathrm{T}}/\sqrt{n}$: normalized MSE. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, and FIC. The best results are in bold face.

n	σ	τ	$\lambda_n = 2$	$\lambda_n = \log(n)pn^{1/2}$	FIC
25	0.5	0.85	0.120	**0.107**	1.000
		0.95	**0.974**	1.938	1.000
	1	0.85	0.248	**0.231**	1.000
		0.95	**0.952**	1.860	1.000
50	0.5	0.85	0.898	**0.755**	1.000
		0.95	1.015	2.473	**1.000**
	1	0.85	**0.951**	1.005	1.000
		0.95	**0.953**	2.387	1.000
100	0.5	0.85	1.003	**0.976**	1.000
		0.95	**0.954**	2.495	1.000
	1	0.85	**0.990**	1.072	1.000
		0.95	**0.959**	2.428	1.000
200	0.5	0.85	**0.980**	0.995	1.000
		0.95	**0.966**	2.950	1.000
	1	0.85	**0.956**	1.073	1.000
		0.95	**0.958**	2.698	1.000

Table S.5: Simulation results of Example I: relative loss. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, and UBERIC$_{PLS}$ with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$. The best results are in bold face.

n	σ	τ	UBERIC $\lambda_n = 2$	UBERIC $\lambda_n = \log(n)pn^{1/2}$	UBERIC$_{PLS}$ $\lambda_n = 2$	UBERIC$_{PLS}$ $\lambda_n = \log(n)pn^{1/2}$
25	0.5	0.85	**1.390**	4.094	1.412	2.853
		0.95	1.412	1.450	1.412	**1.056**
	1	0.85	1.511	5.321	**1.488**	4.514
		0.95	1.465	1.774	1.465	**1.241**
50	0.5	0.85	**1.350**	2.899	**1.350**	1.777
		0.95	1.356	1.069	1.356	**1.028**
	1	0.85	**1.496**	4.971	**1.496**	3.170
		0.95	1.446	1.068	1.446	**1.031**
100	0.5	0.85	**1.261**	1.512	**1.261**	1.049
		0.95	1.192	**1.015**	1.192	**1.015**
	1	0.85	**1.348**	3.417	**1.348**	1.522
		0.95	1.235	**1.008**	1.235	**1.008**
200	0.5	0.85	1.161	**1.006**	1.161	**1.006**
		0.95	1.111	**1.006**	1.111	**1.006**
	1	0.85	1.205	1.553	1.205	**1.008**
		0.95	1.139	**1.003**	1.139	**1.003**

Table S.6: Simulation results of Example II: relative loss. Methods compared are UBERIC with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$, and UBERIC$_{PLS}$ with $\lambda_n = 2$ and $\lambda_n = \log(n)pn^{1/2}$. The best results are in bold face.

			UBERIC	UBERIC	UBERIC$_{PLS}$	UBERIC$_{PLS}$
n	σ	τ	$\lambda_n = 2$	$\lambda_n = \log(n)pn^{1/2}$	$\lambda_n = 2$	$\lambda_n = \log(n)pn^{1/2}$
25	0.5	0.85	0.772	0.576	0.782	**0.792**
		0.95	0.756	0.976	0.756	**0.982**
	1	0.85	0.758	0.468	**0.766**	0.720
		0.95	0.744	0.960	0.744	**0.988**
50	0.5	0.85	0.646	0.832	0.646	**0.954**
		0.95	0.608	0.974	0.608	**0.976**
	1	0.85	0.606	0.696	0.606	**0.904**
		0.95	0.560	0.990	0.560	**0.992**
100	0.5	0.85	0.584	0.962	0.584	**0.978**
		0.95	0.590	**0.982**	0.590	**0.982**
	1	0.85	0.602	0.872	0.602	**0.984**
		0.95	0.614	**0.998**	0.614	**0.998**
200	0.5	0.85	0.548	**0.984**	0.548	**0.984**
		0.95	0.598	**0.996**	0.598	**0.996**
	1	0.85	0.544	0.982	0.544	**1.000**
		0.95	0.584	**1.000**	0.584	**1.000**

References

Claeskens, G. and Hjort, N. L. (2003), "The focused information criterion," *Journal of the American Statistical Association*, 98, 900–916.

Liang, H. and Li, R. (2009), "Variable selection for partially linear models with measurement errors," *Journal of the American Statistical Association*, 104, 234–248.

Magnus, J. R. and Neudecker, H. (1979), "The commutation matrix: Some properties and applications," *Annals of Statistics*, 7, 381–394.

Wang, H. Y., Chen, X., and Flournoy, N. (2016), "The focused information criterion for varying-coefficient partially linear measurement error models," *Statistical Papers*, 57, 99–113.