Supplementary material for

“Multivariate Calibrations with Auxiliary Information”

Pei Feng Tong[1], Song Xi Chen[1,2], and Cheng Yong Tang[3]

[1] *Guanghua School of Management, Peking University, Beijing 100871, China*

[2] *Center for Statistical Science, Peking University, Beijing 100871, China*

[3] *Department of Statistics, Operations and Data Science,*

Temple University, Philadelphia, PA 19122-6083, USA

Supplementary Material

A An Expectation-Maximization algorithm

The model fitting procedure for a mixed model containing random effect can be achieved by the EM algorithm (Dempster et al., 1977), which is an iterated procedure, performing expectation step (E-step) and maximization step (M-step) alternately.

To illustrate the procedure of the EM algorithm, we first assume the covariance matrices $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are known, then with Assumption 3, the joint distribution is

$$\begin{pmatrix} \mathrm{vec}(\boldsymbol{A}^T) \\ \mathrm{vec}(\boldsymbol{Y}^T) \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathrm{vec}(\boldsymbol{\alpha}\mathbf{1}_n^T + \boldsymbol{B}^T\boldsymbol{X}^T + \boldsymbol{D}^T\boldsymbol{Z}^T) \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_k \otimes \boldsymbol{\Gamma}_1 & \boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_1 & \boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 + \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2 \end{pmatrix}\right].$$

Let $\boldsymbol{V} = \mathrm{var}(\mathrm{vec}(\boldsymbol{Y}^T)) = \boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 + \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2$, $\mathcal{B} = [\boldsymbol{\alpha}, \boldsymbol{B}^T, \boldsymbol{D}^T]^T$ and $\mathcal{X} = [\mathbf{1}_n, \boldsymbol{X}, \boldsymbol{Z}]$, the standard estimator of $\mathrm{vec}(\mathcal{B}^T)$ and $\mathrm{vec}(\boldsymbol{A}^T)$ are the GLS estimator

$$\mathrm{vec}(\hat{\mathcal{B}}^T|\boldsymbol{Y},\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = \left[(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{-1}(\mathcal{X} \otimes \boldsymbol{I}_q)\right]^{-1}(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{-1}\mathrm{vec}(\boldsymbol{Y}^T), \tag{A.1}$$

and the posterior mean

$$\mathbb{E}(\mathrm{vec}(\boldsymbol{A}^T)|\boldsymbol{Y},\hat{\mathcal{B}},\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = (\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1)\boldsymbol{V}^{-1}(\mathrm{vec}(\boldsymbol{Y}^T - \hat{\mathcal{B}}^T\mathcal{X}^T)). \tag{A.2}$$

Then, we will discuss the EM algorithm for the estimation of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$.

E-step. In the E-step, we aim to obtain the expectation of the complete log-likelihood, that is,

$$Q(\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2;\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]}) := \mathbb{E}_{\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]}}(-2\log L(\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2)) \tag{A.3}$$

$$=\mathbb{E}_{\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]}}\left[\mathrm{tr}\left\{(\boldsymbol{Y} - \mathcal{X}\mathcal{B}^{[m+1]} - \boldsymbol{\Psi}\boldsymbol{A}^{[m+1]})\boldsymbol{\Gamma}_2^{-1}(\boldsymbol{Y} - \mathcal{X}\mathcal{B}^{[m+1]} - \boldsymbol{\Psi}\boldsymbol{A}^{[m+1]})^T\right\}\right]+$$

$$\mathbb{E}_{\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]}}\left[\mathrm{tr}\left\{\boldsymbol{A}^{[m+1]}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{A}^{[m+1]^T}\right\}\right] + k\log(\det(\boldsymbol{\Gamma}_1)) + n\log(\det(\boldsymbol{\Gamma}_2)) + \mathrm{Const}, \tag{A.4}$$

where we first estimate $\mathcal{B}^{[m+1]}$ according to (A.1) and $\mathrm{vec}(\boldsymbol{A}^{[m+1]^T}) = \mathbb{E}(\mathrm{vec}(\boldsymbol{A}^T)|\boldsymbol{Y},\mathcal{B}^{[m+1]},$ $\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]})$ via (A.2).

M-step. In the M-step, we consider minimizing the negative log-likelihood (A.4). Then, the maximum likelihood estimators of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are

$$\boldsymbol{I}_k \otimes \boldsymbol{\Gamma}_1^{[m+1]} = \boldsymbol{I}_k \otimes \boldsymbol{\Gamma}_1^{[m]} + \mathrm{vec}(\boldsymbol{A}^{[m+1]^T})\mathrm{vec}(\boldsymbol{A}^{[m+1]^T})^T - (\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1^{[m]})\times$$

$$\left\{\boldsymbol{V}^{[m]^{-1}} - \boldsymbol{V}^{[m]^{-1}}(\mathcal{X} \otimes \boldsymbol{I}_q)\left[(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{[m]^{-1}}(\mathcal{X} \otimes \boldsymbol{I}_q)\right]^{-1}(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{[m]^{-1}}\right\}(\boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_1^{[m]}),$$

$$\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2^{[m+1]} = \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2^{[m]} + \mathrm{vec}(\boldsymbol{E}^{[m+1]^T})\mathrm{vec}(\boldsymbol{E}^{[m+1]^T})^T - (\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2^{[m]})\times$$

$$\left\{\boldsymbol{V}^{[m]^{-1}} - \boldsymbol{V}^{[m]^{-1}}(\mathcal{X} \otimes \boldsymbol{I}_q)\left[(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{[m]^{-1}}(\mathcal{X} \otimes \boldsymbol{I}_q)\right]^{-1}(\mathcal{X}^T \otimes \boldsymbol{I}_q)\boldsymbol{V}^{[m]^{-1}}\right\}(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2^{[m]}),$$

where $\mathrm{vec}(\boldsymbol{E}^{[m+1]^T}) = \mathbb{E}(\mathrm{vec}(\boldsymbol{E}^T)|\boldsymbol{Y},\mathcal{B}^{[m+1]},\boldsymbol{\Gamma}_1^{[m]},\boldsymbol{\Gamma}_2^{[m]}) = \mathrm{vec}(\boldsymbol{Y}^T) - \mathrm{vec}(\mathcal{B}^{[m+1]^T}\mathcal{X}^T) - (\boldsymbol{\Psi} \otimes$ $\boldsymbol{I}_q)\mathrm{vec}(\boldsymbol{A}^{[m+1]^T})$. The final $\boldsymbol{\Gamma}_1^{[m+1]}$ and $\boldsymbol{\Gamma}_2^{[m+1]}$ can be calculated by taking the average of the corresponding block diagonal matrices. Here we applied the restricted maximum likelihood

(REML) estimates (Laird and Ware, 1982; Meng and Van Dyk, 1998), correcting the downwards bias of ML variance estimators.

Combining the E-step and M-step, we get a complete iteration procedure. We stop the EM iteration if the change in log-likelihood is small enough.

A.1 Considerations of dimensions

The derivation of EM algorithm involves a calculation of the inverse of a $nq \times nq$ dimensional matrix $\boldsymbol{V}^{-1} = (\boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 + \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2)^{-1}$, which is computational infeasible in practice when n is large. Fortunately, one can simplify the inversion through its special structure. Precisely, we have

$$
\begin{aligned}
(\boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1 + \boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2)^{-1} &= ((\boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1} + \boldsymbol{I}_{nq})(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2))^{-1} \\
&= (\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}_2^{-1})(\boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1} + \boldsymbol{I}_{nq})^{-1}.
\end{aligned}
$$

Then, denote $\boldsymbol{K} = \boldsymbol{\Psi}\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1} + \boldsymbol{I}_{nq} = (\boldsymbol{\Psi} \otimes \boldsymbol{I}_q)(\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1}) + \boldsymbol{I}_{nq}$, using the Sherman–Morrison–Woodbury formula again, we can get

$$
\begin{aligned}
\boldsymbol{K}^{-1} &= \boldsymbol{I}_{nq} - (\boldsymbol{\Psi} \otimes \boldsymbol{I}_q)(\boldsymbol{I}_{kq} + (\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1})(\boldsymbol{\Psi} \otimes \boldsymbol{I}_q))^{-1}(\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1})) \\
&= \boldsymbol{I}_{nq} - (\boldsymbol{\Psi} \otimes \boldsymbol{I}_q)(\boldsymbol{I}_{kq} + \boldsymbol{\Psi}^T\boldsymbol{\Psi} \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1})^{-1}(\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1}).
\end{aligned}
$$

In this expression, the inverse matrix $\boldsymbol{K}^{-1}$ only has dimension $kq \times kq$, thus is fixed and would not increase with the sample size. This is also the reason why we use k basis functions to capture the heterogeneity structure of $\boldsymbol{Y}$. As a result, we can rewrite (A.2) as

$$
\mathrm{vec}(\hat{\boldsymbol{A}}^T) = (\boldsymbol{\Psi}^T \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^{-1})\boldsymbol{K}^{-1}(\mathrm{vec}(\boldsymbol{Y}^T - \hat{\boldsymbol{\alpha}}\boldsymbol{1}_n^T - \hat{\boldsymbol{B}}^T\boldsymbol{X}^T - \hat{\boldsymbol{D}}^T\boldsymbol{Z}^T)). \tag{A.5}
$$

Other equations involve $\boldsymbol{V}^{-1}$ can be simplified analogically.

B Technical development

Based on the Assumption 1, we first establish the GLS estimator.

Lemma B.1. *With Assumption 1, the OLS estimators based on the training sample from (2.3) are* $\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{Y}}$, $\hat{\boldsymbol{B}} = (\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}$ *and* $\hat{\boldsymbol{D}} = (\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Y}$*, while* $\boldsymbol{S} = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}})$ *with* $\hat{\boldsymbol{Y}} = \boldsymbol{1}_n\hat{\boldsymbol{\alpha}}^T + \boldsymbol{X}\hat{\boldsymbol{B}} + \boldsymbol{Z}\hat{\boldsymbol{D}}$*. By plugging the OLS estimators to (2.4) and solving for* $\boldsymbol{\xi}$*, we attain the GLS-type estimator*

$$\hat{\boldsymbol{\xi}}_{gls} = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}').$$

Compared with GLS, the reversed model for deriving (2.7) encounter the endogeneity issue. To see this, we noticed that (2.3) admits an expression

$$\begin{aligned}\boldsymbol{X} =& \boldsymbol{Y}\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1} - \boldsymbol{1}_n\boldsymbol{\alpha}^T\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1} - \boldsymbol{Z}\boldsymbol{D}\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1} - \boldsymbol{E}\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1}\\ :=& \boldsymbol{1}_n\boldsymbol{\theta}^T + \boldsymbol{Y}\boldsymbol{\Phi} + \boldsymbol{Z}\boldsymbol{\Psi} + \boldsymbol{E}_{ir}.\end{aligned}$$

Since $\boldsymbol{Y}$ is correlated with $\boldsymbol{E}$, $\boldsymbol{Y}$ also correlated with the transformed $\boldsymbol{E}_{ir}$, which leads to endogeneity. As we have shown in Theorem 2, the inverse regression estimator is biased to the prior mean of $\boldsymbol{X}$, while the GLS estimator is a unbiased estimator with the true causal direction. The endogeneity of $\hat{\boldsymbol{\xi}}_{ir}$ is not caused by some unobserved confounders, thus can not be handled by the instrumental variable (IV) regression (Greene, 2008).

Moreover, let $c = l^{-1} + n^{-1} + c_3 - c_4 + \hat{\hat{\boldsymbol{y}}}'^T_{res}(\hat{\boldsymbol{Y}}^T_{res}\hat{\boldsymbol{Y}}_{res})^{-1}\hat{\hat{\boldsymbol{y}}}'_{res}$, where $c_4 = \boldsymbol{\zeta}^T\boldsymbol{C}_1\boldsymbol{\zeta}$ and $\hat{\hat{\boldsymbol{y}}}'_{res} = \bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'$, the following theorem derives the Bayes estimator.

Lemma B.2. *With Assumptions 1 and 2, the posterior of* $\boldsymbol{\xi}$

$$P(\boldsymbol{\xi}|\bar{\boldsymbol{y}}', \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') \sim T_{\nu-p+q}\left(\hat{\boldsymbol{\xi}}_{bay}, (\nu - p + q)^{-1}c(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\right)$$

and the Bayes estimator

$$\hat{\boldsymbol{\xi}}_{bay} = \left((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1} + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\hat{\boldsymbol{y}}}'_{res} + \boldsymbol{\zeta}. \tag{B.1}$$

The second part of Lemma B.2 is readily implied from the fact that the mean of the multivariate t-distribution coincides with its mode.

We use a two-stage expression as the intermediate result to derive the equivalence in Theorem 1 (i).

Lemma B.3. *Given Assumption 1, the inverse regression estimator in (2.7) admits an expression*

$$\hat{\boldsymbol{\xi}}_{ir} = \boldsymbol{X}^T \hat{\boldsymbol{Y}}_{res} (\hat{\boldsymbol{Y}}_{res}^T \hat{\boldsymbol{Y}}_{res})^{-1} \hat{\boldsymbol{y}}'_{res} + \boldsymbol{\zeta}, \tag{B.2}$$

where $\hat{\boldsymbol{Y}}_{res} = \boldsymbol{Y} - \boldsymbol{1}_n \hat{\boldsymbol{\alpha}}^T - \mathbb{P}_z \boldsymbol{X} \hat{\boldsymbol{B}} - \boldsymbol{Z} \hat{\boldsymbol{D}}$.

We note that (2.8) and (B.1) are formally equivalent.

Lemma B.4. *With Assumption 1, we have (2.8) and (B.1) are formally equivalent, say*

$$(\boldsymbol{I} + \boldsymbol{H}_{ir})^{-1} \{\hat{\boldsymbol{\xi}}_{gls} - \boldsymbol{\zeta}\} + \boldsymbol{\zeta} = \left((\boldsymbol{X}^T \mathbb{M}_z \boldsymbol{X})^{-1} + \hat{\boldsymbol{B}} \boldsymbol{S}^{-1} \hat{\boldsymbol{B}}^T \right)^{-1} \hat{\boldsymbol{B}} \boldsymbol{S}^{-1} \hat{\boldsymbol{y}}'_{res} + \boldsymbol{\zeta}.$$

The following lemma combines with Lemmas B.3 and B.4 directly lead to Theorem 1 (i).

Lemma B.5. *With Assumption 1, we have*

$$\boldsymbol{X}^T \hat{\boldsymbol{Y}}_{res} (\hat{\boldsymbol{Y}}_{res}^T \hat{\boldsymbol{Y}}_{res})^{-1} \hat{\boldsymbol{y}}'_{res} + \boldsymbol{\zeta} = \left((\boldsymbol{X}^T \mathbb{M}_z \boldsymbol{X})^{-1} + \hat{\boldsymbol{B}} \boldsymbol{S}^{-1} \hat{\boldsymbol{B}}^T \right)^{-1} \hat{\boldsymbol{B}} \boldsymbol{S}^{-1} \hat{\boldsymbol{y}}'_{res} + \boldsymbol{\zeta}.$$

Combining the results from Lemmas B.2 to B.5, we finally obtain Theorem 1 (ii).

C Proof of Lemma B.1

Proof. The GLS solution is a classical estimator, obtained by first regressing $\boldsymbol{Y}$ on $\boldsymbol{X}$ and $\boldsymbol{Z}$ using (2.3), and then solving $\boldsymbol{\xi}$ in (2.4) by the estimated parameters. By plug-in the estimated parameters into equation (2.4), we obtain that

$$\boldsymbol{Y}' = \boldsymbol{1}_l \hat{\boldsymbol{\alpha}}^T + \boldsymbol{1}_l \boldsymbol{\xi}^T \hat{\boldsymbol{B}} + \boldsymbol{Z}' \hat{\boldsymbol{D}} + \boldsymbol{E}',$$

where in this equation, only $\boldsymbol{\xi}$ is unknown. The GLS considers the following optimization problem,

$$\arg\min_{\boldsymbol{\xi}} \operatorname{tr}\left[(\boldsymbol{Y}' - \mathbf{1}_l\hat{\boldsymbol{\alpha}}^T - \mathbf{1}_l\boldsymbol{\xi}^T\hat{\boldsymbol{B}} - \boldsymbol{Z}'\hat{\boldsymbol{D}})\boldsymbol{S}^{-1}(\boldsymbol{Y}' - \mathbf{1}_l\hat{\boldsymbol{\alpha}}^T - \mathbf{1}_l\boldsymbol{\xi}^T\hat{\boldsymbol{B}} - \boldsymbol{Z}'\hat{\boldsymbol{D}})^T\right],$$

which is also equivalent to a weighted least square problem with weighting matrix $\boldsymbol{S}^{-1}$. Using the first order condition, one can obtain the GLS estimator $\hat{\boldsymbol{\xi}}_{gls}$ as described in (2.6). □

D Proof of Lemma B.2

The proof of Lemma B.2 is decomposed into two parts. We first give the proof under the additional condition $\boldsymbol{X}^T\boldsymbol{Z} = 0$, and then generalize the proof without constraint.

We first give the following Lemma D.1, D.2 and D.3, which would be useful for further derivations.

Lemma D.1. *With the model defined by equation* (2.3)*,* (2.4) *together with Assumption 2, we have*

$$P(\boldsymbol{\xi}|\bar{\boldsymbol{y}}', \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') \propto P(\boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')P(\bar{\boldsymbol{y}}'|\boldsymbol{Y}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') := P(\boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')L(\boldsymbol{\xi}). \tag{D.1}$$

Proof. We have the following steps:

$$\begin{aligned}
&P(\boldsymbol{\xi}|\bar{\boldsymbol{y}}', \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') \\
\propto &P(\bar{\boldsymbol{y}}', \boldsymbol{Y}|\boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')P(\boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') \\
= &P(\boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')P(\boldsymbol{Y}|\boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')P(\bar{\boldsymbol{y}}'|\boldsymbol{Y}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}') \\
\propto &P(\boldsymbol{\xi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')P(\bar{\boldsymbol{y}}'|\boldsymbol{Y}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}'),
\end{aligned}$$

which shows the result, where the second equality uses the fact that $\boldsymbol{\xi}$ is independent of $\boldsymbol{X}$. □

Now, the key problem is to construct the distribution of $L(\boldsymbol{\xi})$. In order to do that, we use the noninformative invariant Jefferys prior as stated in Assumption 2 (i). Then, under the condition $\boldsymbol{X}^T\boldsymbol{Z} = 0$, the distribution of $L(\boldsymbol{\xi})$ will be the multivariate t distribution as follows.

Lemma D.2. *Given Assumption 2, denote* $\sigma^2(\boldsymbol{\xi}) = 1/l + 1/n + \boldsymbol{\xi}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\xi} + \bar{\boldsymbol{z}}'^T(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'$ *and* $\hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}) = \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'$, *we have that*

$$L(\boldsymbol{\xi}) \sim T_\nu\left(\hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}), \frac{1}{\nu}\sigma^2(\boldsymbol{\xi})\boldsymbol{S}\right) \propto \frac{(\sigma^2(\boldsymbol{\xi}))^{\frac{\nu}{2}}}{(\sigma^2(\boldsymbol{\xi}) + (\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}))^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi})))^{\frac{\nu+q}{2}}} \tag{D.2}$$

Proof. The proof contains the following steps:

$$(\bar{\boldsymbol{y}}'|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim N(\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \sigma^2(\boldsymbol{\xi})\boldsymbol{\Gamma}), \tag{D.3}$$

$$(\boldsymbol{\Gamma}^{-1}|\boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim W^{-1}(\nu + q - 1, \boldsymbol{S}^{-1}), \tag{D.4}$$

$$(\bar{\boldsymbol{y}}'|\boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim T_\nu(\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \nu^{-1}\sigma^2(\boldsymbol{\xi})\boldsymbol{S}), \tag{D.5}$$

where $W^{-1}(\nu', \boldsymbol{S})$ is the inverse-Wishart distribution with degree of freedom ν'. When $\bar{\boldsymbol{y}}'$ and $\boldsymbol{\Gamma}^{-1}$ satisfy (D.3) and (D.4), we say the joint distribution of $(\bar{\boldsymbol{y}}', \boldsymbol{\Gamma}^{-1})$ has a normal-inverse-Wishart distribution, whose marginal distribution over $\bar{\boldsymbol{y}}'$ is a multivariate t-distribution, leads to (D.5) and the result provides in (D.2).

Proof of (D.3):

We note that $\bar{\boldsymbol{y}}' = \boldsymbol{\alpha} + \boldsymbol{B}^T\boldsymbol{\xi} + \boldsymbol{D}^T\bar{\boldsymbol{z}}' + \bar{\boldsymbol{E}}'$, the proof of (D.3) can be divided into the positerior distribution towards $\boldsymbol{\alpha}$, $\boldsymbol{B}$, $\boldsymbol{D}$ and $\bar{\boldsymbol{E}}'$, respectively.

Since $\boldsymbol{X}$ is of column full rank, we can apply eigenvalue decomposition towards $\boldsymbol{X}^T\boldsymbol{X}/n$. Let $\boldsymbol{P}_1$ be an orthogonal $p \times p$ matrix of eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}/n$, $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_{11}, \ldots, \lambda_{1p})$ be the diagonal matrix of the corresponding eigenvalues, satisfying $\sum_{i=1}^p \lambda_{1i} = p$. We have

$$\boldsymbol{P}_1^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{P}_1 = n\boldsymbol{\Lambda}_1,$$

since $\text{tr}(\boldsymbol{P}_1^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{P}_1) = np = \text{tr}(n\boldsymbol{\Lambda}_1)$ due to $\boldsymbol{X}$ has been standardized. Similarly, we can find a $p' \times p'$ orthogonal matrix $\boldsymbol{P}_2$ such that

$$\boldsymbol{P}_2^T \boldsymbol{Z}^T \boldsymbol{Z} \boldsymbol{P}_2 = n\boldsymbol{\Lambda}_2,$$

with $\boldsymbol{\Lambda}_2 = \text{diag}(\lambda_{21}, \ldots, \lambda_{2p'})$ and $\sum_{i=1}^{p'} \lambda_{2i} = p'$.

As the covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$ is treated as fixed, one can find a $n \times n$ orthogonal matrix $\boldsymbol{Q}$ to get the canonical form so that

$$\boldsymbol{Q}^T \boldsymbol{Y} = \boldsymbol{Q}^T \boldsymbol{1}_n \boldsymbol{\alpha}^T + \boldsymbol{Q}^T \boldsymbol{X}\boldsymbol{B} + \boldsymbol{Q}^T \boldsymbol{Z}\boldsymbol{D} + \boldsymbol{Q}^T \boldsymbol{E} = \begin{pmatrix} n^{1/2}\boldsymbol{\alpha}^T + \boldsymbol{E}_0^T \\ n^{1/2}\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{P}_1^T \boldsymbol{B} + \boldsymbol{E}_1 \\ n^{1/2}\boldsymbol{\Lambda}_2^{1/2}\boldsymbol{P}_2^T \boldsymbol{D} + \boldsymbol{E}_2 \\ \boldsymbol{E}_3 \end{pmatrix} := \begin{pmatrix} \boldsymbol{G}_0^T \\ \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \\ \boldsymbol{G}_3 \end{pmatrix}, \quad \text{(D.6)}$$

where $\boldsymbol{G}_0^T \in \mathbb{R}^{1\times q}$, $\boldsymbol{G}_1 \in \mathbb{R}^{p\times q}$, $\boldsymbol{G}_2 \in \mathbb{R}^{p'\times q}$ and $\boldsymbol{G}_3 \in \mathbb{R}^{(n-p-p'-1)\times q}$. The canonical form can be derived by the singular value decomposition (SVD) of $\mathcal{X} = [\boldsymbol{1}, \boldsymbol{X}, \boldsymbol{Z}]$, such that

$$\mathcal{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{V}^T,$$

where

$$\boldsymbol{\Lambda} = n^{1/2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \boldsymbol{\Lambda}_1^{1/2} & 0 \\ 0 & 0 & \boldsymbol{\Lambda}_2^{1/2} \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \boldsymbol{P}_1 & 0 \\ 0 & 0 & \boldsymbol{P}_2 \end{pmatrix},$$

$\boldsymbol{\Lambda}^T \boldsymbol{\Lambda}$ is diagonal, consists of eigenvalues of $\mathcal{X}^T \mathcal{X}$.

The canonical form of (2.3) is

$$\boldsymbol{Q}^T \boldsymbol{Y} = \boldsymbol{Q}^T \mathcal{X} \begin{pmatrix} \boldsymbol{\alpha}^T \\ \boldsymbol{B} \\ \boldsymbol{D} \end{pmatrix} + \boldsymbol{Q}^T \boldsymbol{E},$$

leading to (D.6).

Since $\boldsymbol{Q}$ is orthogonal and the row vectors of $\boldsymbol{E}$ are IID by $N(0, \boldsymbol{\Gamma})$, the newly generated residuals $\boldsymbol{E}_0^T$, $\boldsymbol{E}_1$, $\boldsymbol{E}_2$ and $\boldsymbol{E}_3$ have IID $N(0, \boldsymbol{\Gamma})$ row vectors, too. To see this, let $\boldsymbol{q}_i$ be the i-th column of $\boldsymbol{Q}$, $\boldsymbol{e}_i^T$ be the i-th row of $\boldsymbol{E}$. Then, the mean and variance of the i-th row of $\boldsymbol{Q}^T\boldsymbol{E}$ are

$$\mu(\boldsymbol{E}^T\boldsymbol{q}_i) = 0 \text{ and } \mathrm{var}(\boldsymbol{E}^T\boldsymbol{q}_i) = \sum_{k=1}^{n} q_{ik}^2 \boldsymbol{\Gamma} = \boldsymbol{\Gamma},$$

and the independence follows from that

$$\mathrm{cov}(\boldsymbol{E}^T\boldsymbol{q}_i, \boldsymbol{E}^T\boldsymbol{q}_j) = \mathrm{cov}(\sum_{k=1}^{n} q_{ik}\boldsymbol{e}_k, \sum_{l=1}^{n} q_{jl}\boldsymbol{e}_l) = \sum_{k=1}^{n} q_{ik}q_{jk}\boldsymbol{\Gamma} = 0, \text{ if } i \neq j.$$

Similarly, we can find a $l \times l$ orthogonal matrix $\boldsymbol{Q}'$ for (2.4) such that

$$\begin{aligned}\boldsymbol{Q}'^T\boldsymbol{Y}' =& \boldsymbol{Q}'^T\boldsymbol{1}_l\boldsymbol{\alpha}^T + \boldsymbol{Q}'^T\boldsymbol{1}_l\boldsymbol{\xi}^T\boldsymbol{B} + \boldsymbol{Q}'^T\boldsymbol{Z}'\boldsymbol{D} + \boldsymbol{Q}'^T\boldsymbol{E}' \\ =& \begin{pmatrix} l^{1/2}(\boldsymbol{\alpha}^T + \boldsymbol{\xi}^T\boldsymbol{B} + \bar{\boldsymbol{z}}'^T\boldsymbol{D}) + \boldsymbol{E}_1'^T \\ \boldsymbol{E}_2' \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{G}_1'^T \\ \boldsymbol{G}_2' \end{pmatrix}.\end{aligned}$$

For example, the first column of $\boldsymbol{Q}'$ is chosen to be $\frac{1}{\sqrt{l}}\boldsymbol{1}_l$, and the following columns are orthogonal towards the unit vector. Indeed, the residuals $\boldsymbol{E}_1'$ and $\boldsymbol{E}_2'$ are IID normal.

The least square estimators of $\boldsymbol{\alpha}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ are given by

$$\hat{\boldsymbol{\alpha}} = n^{-1/2}\boldsymbol{G}_0$$

$$\boldsymbol{P}_1^T\hat{\boldsymbol{B}} = n^{-1/2}\boldsymbol{\Lambda}_1^{-1/2}\boldsymbol{G}_1$$

$$\boldsymbol{P}_2^T\hat{\boldsymbol{D}} = n^{-1/2}\boldsymbol{\Lambda}_2^{-1/2}\boldsymbol{G}_2.$$

Let $\boldsymbol{A}_1 = \boldsymbol{P}_1^T\boldsymbol{B}$, $\boldsymbol{A}_2 = \boldsymbol{P}_2^T\boldsymbol{D}$, $\boldsymbol{a}_{1i}^T$ and $\boldsymbol{a}_{2i}^T$ be the row vectors of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, and $\boldsymbol{G}_{1i}^T$ and $\boldsymbol{G}_{2i}^T$ be the row vector of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$, we further have

$$\hat{\boldsymbol{a}}_{1i} = n^{-1/2}\lambda_{1i}^{-1/2}\boldsymbol{G}_{1i}, i = 1, \ldots, p$$

$$\hat{\boldsymbol{a}}_{2i} = n^{-1/2}\lambda_{2i}^{-1/2}\boldsymbol{G}_{2i}, i = 1, \ldots, p'.$$

Since $\boldsymbol{G}_0 \sim \mathcal{N}(n^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Gamma})$, $\boldsymbol{G}_{1i} \sim \mathcal{N}(n^{1/2}\lambda_{1i}^{1/2}\boldsymbol{a}_{1i}, \boldsymbol{\Gamma})$, $\boldsymbol{G}_{2i} \sim \mathcal{N}(n^{1/2}\lambda_{2i}^{1/2}\boldsymbol{a}_{2i}, \boldsymbol{\Gamma})$ and they are independent, the distribution of LS estimators are $\hat{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{\alpha}, n^{-1}\boldsymbol{\Gamma})$, $\hat{\boldsymbol{a}}_{1i} \sim \mathcal{N}(\boldsymbol{a}_{1i}, n^{-1}\lambda_{1i}^{-1}\boldsymbol{\Gamma})$, $\hat{\boldsymbol{a}}_{2i} \sim \mathcal{N}(\boldsymbol{a}_{2i}, n^{-1}\lambda_{2i}^{-1}\boldsymbol{\Gamma})$, while $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{a}}_{11}$, ..., $\hat{\boldsymbol{a}}_{1p}$, $\hat{\boldsymbol{a}}_{21}$, ..., $\hat{\boldsymbol{a}}_{2p'}$ are independent.

To obtain the posterior distribution of $\boldsymbol{B}$, we note that condition on $\{\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Gamma}\}$, $\{\boldsymbol{\xi}, \boldsymbol{Z}'\}$ and $\{\boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{D}\}$ are independent. Recall that $\mathcal{X} = [\boldsymbol{1}, \boldsymbol{X}, \boldsymbol{Z}]$, let $\mathcal{B} = [\boldsymbol{\alpha}, \boldsymbol{B}^T, \boldsymbol{D}^T]^T$, the full log likelihood associated with $\mathcal{B}$ is propotional to

$$\begin{aligned}
& -\frac{1}{2}\mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\mathcal{B})^T) \\
=& -\frac{1}{2}\mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}} + \mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}} + \mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T) \\
=& -\frac{1}{2}[\mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})^T) + \mathrm{tr}((\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T) + \\
& \mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})\boldsymbol{\Gamma}^{-1}(\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T) + \mathrm{tr}((\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})^T)] \\
=& -\frac{1}{2}[\mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})^T) + \mathrm{tr}((\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T) + \\
& \mathrm{tr}(\mathcal{X}^T(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})\boldsymbol{\Gamma}^{-1}(\hat{\mathcal{B}} - \mathcal{B})^T) + \mathrm{tr}((\hat{\mathcal{B}} - \mathcal{B})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})^T\mathcal{X})] \\
=& -\frac{1}{2}[\mathrm{tr}((\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}})^T) + \mathrm{tr}((\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T)]. \qquad \text{(D.7)}
\end{aligned}$$

The last equality is due to $\mathcal{X}^T(\boldsymbol{Y} - \mathcal{X}\hat{\mathcal{B}}) = 0$ in OLS.

Moreover, condition on $\{\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}'\}$, $\hat{\mathcal{B}}$ provides no additional information. Thus, $(\mathcal{B}|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') = (\mathcal{B}|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}', \hat{\mathcal{B}})$, and with the noninformative invariant Jefferys prior, the posterior distribution

$$\begin{aligned}
\log P(\mathcal{B}|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}', \hat{\mathcal{B}}) \propto & \mathrm{tr}((\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})\boldsymbol{\Gamma}^{-1}(\mathcal{X}\hat{\mathcal{B}} - \mathcal{X}\mathcal{B})^T) \\
=& \mathrm{tr}(\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{V}^T(\hat{\mathcal{B}} - \mathcal{B})\boldsymbol{\Gamma}^{-1}(\hat{\mathcal{B}} - \mathcal{B})^T\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{Q}^T) \\
=& \mathrm{tr}(\boldsymbol{\Lambda}^2(\hat{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{\Gamma}^{-1}(\hat{\boldsymbol{A}} - \boldsymbol{A})^T), \qquad \text{(D.8)}
\end{aligned}$$

where $\boldsymbol{A} = [\boldsymbol{\alpha}, \boldsymbol{A}_1^T, \boldsymbol{A}_2^T]^T$. One can easily check that this is a Gaussian core function.

As a result, we have that

$$
\begin{aligned}
&(\boldsymbol{\alpha}|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \sim \mathcal{N}(\hat{\boldsymbol{\alpha}}, n^{-1}\boldsymbol{\Gamma}),\\
&(\boldsymbol{B}^T\boldsymbol{\xi}|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \sim \mathcal{N}(\hat{\boldsymbol{B}}^T\boldsymbol{\xi}, (n^{-1}\sum_{i=1}^{p}(\boldsymbol{\xi}^T\boldsymbol{P}_{1i})^2/\lambda_{1i})\boldsymbol{\Gamma}),\\
&(\boldsymbol{D}^T\bar{\boldsymbol{z}}'|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \sim \mathcal{N}(\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', n^{-1}(\sum_{i=1}^{p'}(\bar{\boldsymbol{z}}'^T\boldsymbol{P}_{2i})^2/\lambda_{2i})\boldsymbol{\Gamma}),\\
&(\bar{\boldsymbol{E}}'|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \sim \mathcal{N}(0, l^{-1}\boldsymbol{\Gamma}),
\end{aligned}
$$

where $\boldsymbol{P}_{1i}$ and $\boldsymbol{P}_{2i}$ are the column vectors of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, respectively. Since these four conditional distributions are independent, we can derive that

$$
\begin{aligned}
&(\bar{\boldsymbol{y}}'|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}')\\
\sim&\mathcal{N}\left(\hat{\boldsymbol{\alpha}}+\hat{\boldsymbol{B}}^T\boldsymbol{\xi}+\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \left\{l^{-1}+n^{-1}+\boldsymbol{\xi}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\xi}+\bar{\boldsymbol{z}}'^T(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\right\}\boldsymbol{\Gamma}\right),
\end{aligned}
$$

follows from $\bar{\boldsymbol{y}}' = \boldsymbol{\alpha}+\boldsymbol{B}^T\boldsymbol{\xi}+\boldsymbol{D}^T\bar{\boldsymbol{z}}'+\bar{\boldsymbol{E}}'$ and

$$
\frac{1}{n}\sum_{i=1}^{p}(\boldsymbol{\xi}^T\boldsymbol{P}_{1i})^2/\lambda_{1i} = \frac{1}{n}\boldsymbol{\xi}^T\boldsymbol{P}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{P}_1^T\boldsymbol{\xi} = \frac{1}{n}\boldsymbol{\xi}^T(\boldsymbol{P}_1\boldsymbol{\Lambda}_1\boldsymbol{P}_1^T)^{-1}\boldsymbol{\xi} = \boldsymbol{\xi}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{\xi}.
$$

Proof of (D.4): It is well known that the sample covariance matrix $(\boldsymbol{S}|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \sim W(\nu+q-1,\boldsymbol{\Gamma})$, that is,

$$
P(\boldsymbol{S}|\boldsymbol{\Gamma},\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') = \frac{|\boldsymbol{S}|^{\frac{\nu-2}{2}}\exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}^{-1}\boldsymbol{S}))}{2^{\frac{(\nu+q-1)q}{2}}|\boldsymbol{\Gamma}|^{\frac{\nu+q-1}{2}}\Gamma_q(\frac{\nu+q-1}{2})},
$$

and $\Gamma_q(\cdot)$ is a multivariate gamma function. Similar to (D.8), condition on $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$, $\boldsymbol{S}$ provide no more information. Given the noninformative invariant Jefferys prior from Assumption 2(i), say, $P(\boldsymbol{\alpha},\boldsymbol{B},\boldsymbol{D},\boldsymbol{\Gamma}) \propto |\boldsymbol{\Gamma}|^{-(q+1)/2}$, we have the corresponding posterior distribution

$$
\begin{aligned}
P(\boldsymbol{\Gamma}|\boldsymbol{\xi},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{Z}') \propto&|\boldsymbol{\Gamma}|^{-\frac{\nu+q-1}{2}}\exp(-\mathrm{tr}(\boldsymbol{\Gamma}^{-1}\boldsymbol{S})/2)\times|\boldsymbol{\Gamma}|^{-(q+1)/2}\\
=&|\boldsymbol{\Gamma}|^{-\frac{\nu+2q}{2}}\exp(-\mathrm{tr}(\boldsymbol{\Gamma}^{-1}\boldsymbol{S})/2),
\end{aligned}
$$

which is the core function of (D.4).

Proof of (D.5): Following (D.3), (D.4), the joint distribution of $(\bar{\boldsymbol{Y}}', \boldsymbol{\Gamma}^{-1})$ will have a normal-inverse-Wishart distribution, whose marginal distribution over $\bar{\boldsymbol{Y}}'$ is a multivariate-t distribution (Gelman et al., 2013) (p. 73). To be precise, we obtain

$$(\bar{\boldsymbol{y}}'|\boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim T_\nu(\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \nu^{-1}\sigma^2(\boldsymbol{\xi})\boldsymbol{S}).$$

□

Now we are ready to establish the theorem without the condition $\boldsymbol{X}^T\boldsymbol{Z} = 0$.

Lemma D.3. *Given Assumption 2, denote* $\sigma^2(\boldsymbol{\xi}) = 1/l + 1/n + \boldsymbol{\xi}^T\boldsymbol{C}_1\boldsymbol{\xi} - \boldsymbol{c}_2^T\boldsymbol{\xi} + c_3$*, where*

$$\boldsymbol{C}_1 = (\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}$$

$$\boldsymbol{c}_2 = ((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}'$$

$$c_3 = \bar{\boldsymbol{z}}'^T(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}',$$

are the same as what in Assumption 2. Further let $\hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}) = \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'$*, we have that*

$$L(\boldsymbol{\xi}) \sim T_\nu\left(\hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}), \frac{1}{\nu}\sigma^2(\boldsymbol{\xi})\boldsymbol{S}\right) \propto \frac{(\sigma^2(\boldsymbol{\xi}))^{\frac{\nu}{2}}}{(\sigma^2(\boldsymbol{\xi}) + (\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi}))^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}'(\boldsymbol{\xi})))^{\frac{\nu+q}{2}}} \tag{D.9}$$

Proof. The derivations of (D.4) and (D.5) do not require the orthogonal assumption. We only need to show that

$$(\bar{\boldsymbol{y}}'|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim N(\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \sigma^2(\boldsymbol{\xi})\boldsymbol{\Gamma}) \tag{D.10}$$

holds when $\boldsymbol{Z}^T\boldsymbol{X} \neq 0$.

Denote $\boldsymbol{W} = [\boldsymbol{X}, \boldsymbol{Z}]$ and then apply eigenvalue decomposition towards $\boldsymbol{W}^T\boldsymbol{W}$. Let $\boldsymbol{P}$ be an orthogonal $(p+p') \times (p+p')$ matrix of eigenvectors of $\boldsymbol{W}^T\boldsymbol{W}$, $\boldsymbol{\Lambda}_w = \text{diag}(\lambda_1, \ldots, \lambda_{p+p'})$ be the diagonal matrix of corresponding eigenvalues, satisfying $\sum_{i=1}^{p+p'}\lambda_i = p + p'$. Then we have

$$\boldsymbol{P}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{P} = n\boldsymbol{\Lambda}_w,$$

noting that $\mathrm{tr}(\boldsymbol{P}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{P}) = n(p+p') = \mathrm{tr}(n\Lambda_w)$ due to $\boldsymbol{W}$ has been standardized.

Also let $\boldsymbol{\Theta} = [\boldsymbol{B}^T, \boldsymbol{D}^T]^T$ be the stacked parameters. As the covariate $\boldsymbol{X}$ is treated as fixed, one can find a $n \times n$ orthogonal matrix $\boldsymbol{Q}$ to get the canonical form so that

$$\boldsymbol{Q}^T\boldsymbol{Y} = \boldsymbol{Q}^T\mathbf{1}_n\boldsymbol{\alpha}^T + \boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{B} + \boldsymbol{Q}^T\boldsymbol{Z}\boldsymbol{D} + \boldsymbol{Q}^T\boldsymbol{E} = \begin{pmatrix} n^{1/2}\boldsymbol{\alpha}^T + \boldsymbol{E}_0^T \\ n^{1/2}\boldsymbol{\Lambda}_w^{1/2}\boldsymbol{P}^T\boldsymbol{\Theta} + \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{G}_0^T \\ \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \end{pmatrix}, \quad \text{(D.11)}$$

where $\boldsymbol{G}_0^T \in \mathbb{R}^{1\times q}$, $\boldsymbol{G}_1 \in \mathbb{R}^{(p+p')\times q}$ and $\boldsymbol{G}_2 \in \mathbb{R}^{(n-p-p'-1)\times q}$. The canonical form can be derived by the singular value decomposition (SVD) of $[\mathbf{1}, \boldsymbol{W}]$, such that

$$[\mathbf{1}, \boldsymbol{W}] = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{V}^T,$$

where

$$\boldsymbol{\Lambda} = n^{1/2}\begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\Lambda}_w^{1/2} \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{P} \end{pmatrix},$$

$\boldsymbol{\Lambda}^T\boldsymbol{\Lambda}$ is diagonal, consists of eigenvalues of $[\mathbf{1}, \boldsymbol{W}]^T[\mathbf{1}, \boldsymbol{W}]$. The canonical form of (2.3) is

$$\boldsymbol{Q}^T\boldsymbol{Y} = \boldsymbol{Q}^T[\mathbf{1}, \boldsymbol{W}]\begin{pmatrix} \boldsymbol{\alpha}^T \\ \boldsymbol{B} \\ \boldsymbol{D} \end{pmatrix} + \boldsymbol{Q}^T\boldsymbol{E},$$

leading to (D.11).

Since $\boldsymbol{Q}$ is orthogonal and the row vectors of $\boldsymbol{E}$ are IID by $N(0, \boldsymbol{\Gamma})$, the newly generated residuals $\boldsymbol{E}_0^T$, $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ have IID $N(0, \boldsymbol{\Gamma})$ row vectors, too. Similarly, we can find a $l \times l$

orthogonal matrix $\boldsymbol{Q}'$ such that

$$\begin{aligned}\boldsymbol{Q}'^T\boldsymbol{Y}' =&\boldsymbol{Q}'^T\boldsymbol{1}_l\boldsymbol{\alpha}^T + \boldsymbol{Q}'^T\boldsymbol{1}_l\boldsymbol{\xi}^T\boldsymbol{B} + \boldsymbol{Q}'^T\boldsymbol{1}_l\boldsymbol{Z}'^T\boldsymbol{D} + \boldsymbol{Q}'^T\boldsymbol{E}' \\ =& \begin{pmatrix} l^{1/2}(\boldsymbol{\alpha}^T + \boldsymbol{\xi}^T\boldsymbol{B} + \boldsymbol{Z}'^T\boldsymbol{D}) + \boldsymbol{E}_1'^T \\ \boldsymbol{E}_2' \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{G}_1'^T \\ \boldsymbol{G}_2' \end{pmatrix}.\end{aligned}$$

For example, the first column of $\boldsymbol{Q}'$ is chosen to be $\frac{1}{\sqrt{l}}\boldsymbol{1}_l$, and the following columns are orthogonal towards the unit vector. Indeed, the residuals $\boldsymbol{e}_1'$ and $\boldsymbol{e}_2'$ are IID normal.

The least square estimators of $\boldsymbol{\alpha}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ are given by

$$\hat{\boldsymbol{\alpha}} = n^{-1/2}\boldsymbol{G}_0$$

$$\boldsymbol{P}^T\hat{\boldsymbol{\Theta}} = n^{-1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{G}_1$$

Let $\boldsymbol{A} = \boldsymbol{P}^T\boldsymbol{\Theta}$, $\boldsymbol{a}_i^T$ be the row vector of $\boldsymbol{A}$, and $\boldsymbol{G}_i^T$ be the row vector of $\boldsymbol{G}$, we further have

$$\hat{\boldsymbol{a}}_i = n^{-1/2}\lambda_i^{-1/2}\boldsymbol{G}_i, i = 1, \ldots, p + p'$$

Since $\boldsymbol{G}_0 \sim \mathcal{N}(n^{1/2}\boldsymbol{\alpha}, \boldsymbol{\Gamma})$, $\boldsymbol{G}_i \sim \mathcal{N}(n^{1/2}\lambda_i^{1/2}\boldsymbol{a}_i, \boldsymbol{\Gamma})$ and they are independent, the distribution of OLS estimators are $\hat{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{\alpha}, n^{-1}\boldsymbol{\Gamma})$, $\hat{\boldsymbol{a}}_i \sim \mathcal{N}(\boldsymbol{a}_i, n^{-1}\lambda_i^{-1}\boldsymbol{\Gamma})$, while $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{a}}_1$, ..., $\hat{\boldsymbol{a}}_{p+p'}$ are independent.

The posterior distribution is derived via the same discussions in (D.7) and (D.8). As a result, let $\boldsymbol{w}' = [\boldsymbol{\xi}^T, \boldsymbol{Z}'^T]^T$, we have that $P(\boldsymbol{\alpha}+\boldsymbol{B}^T\boldsymbol{\xi}+\boldsymbol{D}^T\boldsymbol{Z}'|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim \mathcal{N}(\hat{\boldsymbol{\alpha}}+\hat{\boldsymbol{B}}^T\boldsymbol{\xi}+\hat{\boldsymbol{D}}^T\boldsymbol{Z}', n^{-1}\boldsymbol{\Gamma}(1 + \sum_{i=1}^{p+p'}(\boldsymbol{w}'^T\boldsymbol{P}_i)^2/\lambda_i))$, where $\boldsymbol{P}_i$ are the column vectors of $\boldsymbol{P}$. Moreover, we can derive that

$$(\bar{\boldsymbol{y}}'|\boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}') \sim \mathcal{N}\left(\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{B}}^T\boldsymbol{\xi} + \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}', \left\{\frac{1}{l} + \frac{1}{n} + \boldsymbol{w}'^T(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{w}'\right\}\boldsymbol{\Gamma}\right),$$

follows from $\bar{\boldsymbol{y}}' = \boldsymbol{\alpha} + \boldsymbol{B}^T\boldsymbol{\xi} + \boldsymbol{D}^T\bar{\boldsymbol{z}}' + \bar{\boldsymbol{E}}'$ and

$$\frac{1}{n}\sum_{i=1}^{p}(\boldsymbol{w}'^T\boldsymbol{P}_i)^2/\lambda_i = \frac{1}{n}\boldsymbol{w}'^T\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^T\boldsymbol{w}' = \frac{1}{n}\boldsymbol{w}'^T(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T)^{-1}\boldsymbol{w}' = \boldsymbol{w}'^T(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{w}'.$$

Adopting the matrix inversion of $(\boldsymbol{W}^T\boldsymbol{W})^{-1}$, one can verify that $\boldsymbol{w}'^T(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{w}' = \boldsymbol{\xi}^T\boldsymbol{C}_1\boldsymbol{\xi} - \boldsymbol{c}_2^T\boldsymbol{\xi} + c_3$, and we have the result in equation (D.10). □

Proof of Lemma B.2.

Proof. First, from Assumption 2 (ii) and the definition of multivariate t-distribution, we note that the prior $P(\boldsymbol{\xi}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Z}') \propto (\sigma^2(\boldsymbol{\xi}))^{-\frac{\nu}{2}}$. Then from Lemma D.3, we see this prior will exactly cancel the numerator of (D.9), which leads to

$$P(\boldsymbol{\xi}|\bar{\boldsymbol{y}}',\boldsymbol{Y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Z}') = P(\boldsymbol{\xi}|\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Z}')L(\boldsymbol{\xi}) \propto \Big(\sigma^2(\boldsymbol{\xi}) + (\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}')^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}')\Big)^{-\frac{\nu+q}{2}}.$$

Then, we will show that the remaining denominator of (D.9) is the core function of the multivariate t-distribution as shown in Lemma B.2. To see this,

$$\begin{aligned}
&\Big(\sigma^2(\boldsymbol{\xi}) + (\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}')^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\bar{\boldsymbol{y}}}')\Big)^{-\frac{\nu+q}{2}}\\
=&\Big(\boldsymbol{\xi}^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)\boldsymbol{\xi} - (\boldsymbol{c}_2^T + 2(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)\boldsymbol{\xi} +\\
&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') + l^{-1} + n^{-1} + c_3\Big)^{-\frac{\nu+q}{2}}\\
=&\Big((\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay})^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay}) - \hat{\boldsymbol{\xi}}_{bay}^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)\hat{\boldsymbol{\xi}}_{bay} +\\
&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') + l^{-1} + n^{-1} + c_3\Big)^{-\frac{\nu+q}{2}}.
\end{aligned}$$

In addition, we noticed that

$$\begin{aligned}
&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - \hat{\boldsymbol{\xi}}_{bay}^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)\hat{\boldsymbol{\xi}}_{bay}\\
=&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T(\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1})\times\\
&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - \boldsymbol{\zeta}^T\boldsymbol{C}_1\boldsymbol{\zeta}\\
=&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T(\boldsymbol{S} + \hat{\boldsymbol{B}}^T\boldsymbol{C}_1^{-1}\hat{\boldsymbol{B}})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - c_4\\
=&(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T(\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - c_4,
\end{aligned}$$

where the second equation is from the SMW formula, the last equation uses the fact that $\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res} = (\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^T - \mathbb{P}_z\boldsymbol{X}\hat{\boldsymbol{B}} - \boldsymbol{Z}\hat{\boldsymbol{D}})^T(\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^T - \mathbb{P}_z\boldsymbol{X}\hat{\boldsymbol{B}} - \boldsymbol{Z}\hat{\boldsymbol{D}}) = \hat{\boldsymbol{B}}^T\boldsymbol{C}_1^{-1}\hat{\boldsymbol{B}} + \boldsymbol{S}$.

Finally,

$$
\begin{aligned}
&P(\boldsymbol{\xi}|\bar{\boldsymbol{y}}', \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Z}')\\
\propto&\Big((\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay})^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay}) + (\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')^T\times\\
&(\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') + l^{-1} + n^{-1} + c_3 - c_4\Big)^{-\frac{\nu+q}{2}}\\
=&((\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay})^T(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}_{bay}) + c)^{-\frac{\nu+q}{2}},
\end{aligned}
$$

which is the core function of $T_{\nu-p+q}\left(\hat{\boldsymbol{\xi}}_{bay}, \frac{1}{\nu-p+q}c(\boldsymbol{C}_1 + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\right)$. □

E Proof of Lemma B.3

To obtain the explicit formula of (2.7), recall that the inverse regression estimator is built under the reversed models $\boldsymbol{X} = \mathbf{1}_n\boldsymbol{\theta}^T + \boldsymbol{Y}\boldsymbol{\Phi} + \boldsymbol{Z}\boldsymbol{\Psi} + \tilde{\boldsymbol{E}}$, $\boldsymbol{\xi} = \boldsymbol{\theta} + \boldsymbol{\Phi}^T\bar{\boldsymbol{y}}' + \boldsymbol{\Psi}^T\bar{\boldsymbol{z}}' + \bar{\boldsymbol{E}}_{ir}'$. Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}, \boldsymbol{\Phi}^T, \boldsymbol{\Psi}^T]^T$ and $\mathcal{Y} = [\mathbf{1}_n, \boldsymbol{Y}, \boldsymbol{Z}]$, based on the first model, the OLS estimators for the parameters are $\hat{\boldsymbol{\Theta}} = (\mathcal{Y}^T\mathcal{Y})^{-1}\mathcal{Y}^T\boldsymbol{X}$. The specific form of $\boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$ are $\hat{\boldsymbol{\theta}}^T = -(\mathbf{1}_n^T(\boldsymbol{I} - \boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\mathbf{1}_n)^{-1}\mathbf{1}_n^T\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{X}$, $\hat{\boldsymbol{\Phi}} = (\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\tilde{\boldsymbol{Y}}\mathbb{M}_z\boldsymbol{X}^T$ and $\hat{\boldsymbol{\Psi}} = (\boldsymbol{Z}^T\mathbb{M}_{\tilde{y}}\boldsymbol{Z})^{-1}\boldsymbol{Z}\mathbb{M}_{\tilde{y}}\boldsymbol{X}$, respectively, where $\tilde{\boldsymbol{Y}} = \mathbb{M}_1\boldsymbol{Y}$, $\mathbb{M}_1 = \boldsymbol{I} - \mathbf{1}_n(\mathbf{1}_n^T\mathbf{1}_n)^{-1}\mathbf{1}_n^T$, $\mathbb{M}_z = \boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$ and $\mathbb{M}_{\tilde{y}} = \boldsymbol{I} - \tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{Y}})^{-1}\tilde{\boldsymbol{Y}}^T$. Substituting them to the second inverse regression model above to attain the inverse regression estimator

$$\hat{\boldsymbol{\xi}}_{ir} = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\Phi}}^T\bar{\boldsymbol{y}}' + \hat{\boldsymbol{\Psi}}^T\bar{\boldsymbol{z}}'$$

Similar as before, we first give the proof of Lemma B.3 with the condition $\boldsymbol{X}^T\boldsymbol{Z} = 0$ as follows:

Lemma E.1. *Given Assumption 1, when $\boldsymbol{X}^T\boldsymbol{Z} = 0$, the inverse regression estimator* (2.7) *admits an expression*

$$\hat{\boldsymbol{\xi}}_{ir} = \boldsymbol{X}^T \hat{\boldsymbol{Y}}_{res} (\hat{\boldsymbol{Y}}_{res}^T \hat{\boldsymbol{Y}}_{res})^{-1} (\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T \bar{\boldsymbol{z}}'), \tag{E.1}$$

where $\hat{\boldsymbol{Y}}_{res} = \boldsymbol{Y} - \mathbf{1}_n \hat{\boldsymbol{\alpha}}^T - \boldsymbol{Z}\hat{\boldsymbol{D}}$.

Proof. The proof contains two parts. We first give the equivalence result by assuming that the mean of $\boldsymbol{Y}$ is 0, and then we give the full proof with the intercept involved.

(i). When $\bar{\boldsymbol{y}} = 0$, then we have $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}} = 0$ and $\tilde{\boldsymbol{Y}} = \boldsymbol{Y}$. Since $\hat{\boldsymbol{D}} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{Y}$, $\hat{\boldsymbol{Y}}_{res} = \boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{D}} = \mathbb{M}_z \boldsymbol{Y}$, our goal is to proof that

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{ir} =& \boldsymbol{X}^T \mathbb{M}_z \boldsymbol{Y} (\boldsymbol{Y}^T \mathbb{M}_z \boldsymbol{Y})^{-1} \bar{\boldsymbol{y}}' + \boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z} (\boldsymbol{Z}^T \mathbb{M}_y \boldsymbol{Z})^{-1} \bar{\boldsymbol{z}}' \\ =& \boldsymbol{X}^T \hat{\boldsymbol{Y}}_{res} (\hat{\boldsymbol{Y}}_{res}^T \hat{\boldsymbol{Y}}_{res})^{-1} (\bar{\boldsymbol{y}}' - \hat{\boldsymbol{D}}^T \bar{\boldsymbol{z}}'),\end{aligned}$$

or

$$\boldsymbol{X}^T \mathbb{M}_z \boldsymbol{Y} (\boldsymbol{Y}^T \mathbb{M}_z \boldsymbol{Y})^{-1} \boldsymbol{Y}^T \boldsymbol{Z} (\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \bar{\boldsymbol{z}}' = -\boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z} (\boldsymbol{Z}^T \mathbb{M}_y \boldsymbol{Z})^{-1} \bar{\boldsymbol{z}}'. \tag{E.2}$$

The right-hand side of (E.2) can be denoted as

$$\begin{aligned}& -\boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z} (\boldsymbol{Z}^T \mathbb{M}_y \boldsymbol{Z})^{-1} \bar{\boldsymbol{z}}' \\ =& -\boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z} ((\boldsymbol{Z}^T\boldsymbol{Z})^{-1} + (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}' \\ =& -\boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z} ((\boldsymbol{Z}^T\boldsymbol{Z})^{-1} + (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}' \\ =& -\boldsymbol{X}^T \mathbb{M}_y (\boldsymbol{I} + \mathbb{P}_z \boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\ =& -\boldsymbol{X}^T \mathbb{M}_y (\boldsymbol{I} - \mathbb{M}_z \boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\ =& -\boldsymbol{X}^T \mathbb{M}_y \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' + \boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' - \boldsymbol{X}^T\mathbb{P}_y\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\ =& -\boldsymbol{\zeta} + \boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\ =& \boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}',\end{aligned}$$

where $\mathbb{P}_z = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$ and the first equivalence used the Sherman–Morrison–Woodbury (SMW) formula. Since it is equivalent to the left hand side of (E.2), thus we finished our proof without intercept.

(ii). To study the effect of the intercept, recall that $\tilde{\boldsymbol{Y}} = \mathbb{M}_1\boldsymbol{Y}$ and $\mathbb{P}_1 = \boldsymbol{1}_n(\boldsymbol{1}_n^T\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^T$, we first give some useful results as below:

$$
\begin{aligned}
\mathbb{M}_z\mathbb{P}_1 =&\mathbb{P}_1\\
\hat{\boldsymbol{Y}}_{res} =&\boldsymbol{Y} - \boldsymbol{1}_n\hat{\boldsymbol{\alpha}}^T - \boldsymbol{Z}\hat{\boldsymbol{D}} = \mathbb{M}_z\boldsymbol{Y} - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Y} = \mathbb{M}_z\boldsymbol{Y} - \boldsymbol{1}_n(\boldsymbol{1}_n^T\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^T\boldsymbol{Y}\\
=&\mathbb{M}_z\boldsymbol{Y} - \mathbb{P}_1\boldsymbol{Y} = \mathbb{M}_z\boldsymbol{Y} - \mathbb{M}_z\mathbb{P}_1\boldsymbol{Y} = \mathbb{M}_z\tilde{\boldsymbol{Y}}
\end{aligned}
$$

Then, our goal is to prove that

$$
\begin{aligned}
&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n(\boldsymbol{1}_n^T(\boldsymbol{I} - \boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\boldsymbol{1}_n)^{-1}+\\
&\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}' + \boldsymbol{X}^T\mathbb{M}_{\tilde{y}}\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_{\tilde{y}}\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}').
\end{aligned}
$$

We note the equivalence in part (i) still holds, say, $\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}' +$ $\boldsymbol{X}^T\mathbb{M}_{\tilde{y}}\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_{\tilde{y}}\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' = \boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')$. We only need to show that

$$
\begin{aligned}
&\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n(\boldsymbol{1}_n^T(\boldsymbol{I} - \boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\boldsymbol{1}_n)^{-1}\\
=&\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}^T\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}. \qquad \text{(E.3)}
\end{aligned}
$$

Denote $\boldsymbol{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n = c_y$, using the SMW formula,

$$
\begin{aligned}
&(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1} = ((\tilde{\boldsymbol{Y}} + \mathbb{P}_1\boldsymbol{Y})^T\mathbb{M}_z(\tilde{\boldsymbol{Y}} + \mathbb{P}_1\boldsymbol{Y}))^{-1}\\
=&(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1} - (\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n(\boldsymbol{1}_n^T\boldsymbol{1}_n + \boldsymbol{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\\
=&(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1} - \frac{1}{n + c_y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\boldsymbol{1}_n\boldsymbol{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
(\mathbf{1}_n^T(\boldsymbol{I}-\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T)\mathbf{1}_n)^{-1} =& \frac{1}{n}-\frac{1}{n}\mathbf{1}_n^T\boldsymbol{Y}(-\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y}+\boldsymbol{Y}^T\mathbb{P}_1\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\mathbf{1}_n\frac{1}{n}\\
=& \frac{1}{n}+\frac{1}{n}\mathbf{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\mathbf{1}_n\frac{1}{n}\\
=& \frac{1}{n}(1+\frac{c_y}{n})
\end{aligned}
$$

Using these results, the left hand side of (E.3) is

$$
\begin{aligned}
\text{LHS} =& \boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}\Big\{(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}-\frac{1}{n+c_y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\Big\}\bar{\boldsymbol{y}}(1+\frac{c_y}{n})\\
=& \boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}(1+\frac{c_y}{n})-\frac{1}{n}\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\boldsymbol{Y}^T\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{Y}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}\\
=& \boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}(1+\frac{c_y}{n})-\frac{c_y}{n}\boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}\\
=& \boldsymbol{X}^T\mathbb{M}_z\tilde{\boldsymbol{Y}}(\tilde{\boldsymbol{Y}}\mathbb{M}_z\tilde{\boldsymbol{Y}})^{-1}\bar{\boldsymbol{y}}=\text{RHS}
\end{aligned}
$$

Thus, we have finished our proof. □

Proof of Lemma B.3.

For the general case without the additional constraint $\boldsymbol{X}^T\boldsymbol{Z}=0$, we follow the similar proof steps as illustrated in Lemma E.1

Proof. Without loss of generality, we assume that the mean of $\boldsymbol{Y}$ is 0, that is to say $\hat{\boldsymbol{\alpha}}=0$, and we aim to prove:

$$
\begin{aligned}
&\boldsymbol{X}^T\hat{\boldsymbol{Y}}_{res}(\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res})^{-1}(\bar{\boldsymbol{y}}'-\hat{\boldsymbol{B}}^T\boldsymbol{\zeta}-\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')+\boldsymbol{\zeta}\\
=&\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\bar{\boldsymbol{y}}'+\boldsymbol{X}^T\mathbb{M}_y\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_y\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \qquad \text{(E.4)}
\end{aligned}
$$

Denote $\boldsymbol{L}_z^y=\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_y\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$ for simplicity and $\boldsymbol{L}_x^z,\boldsymbol{L}_z^x,\boldsymbol{L}_z^y$ similarly. From definition, we can directly get $\boldsymbol{L}_z^y\mathbb{M}_y\mathbb{P}_z=\mathbb{P}_z$ and $\boldsymbol{L}_z^y\mathbb{P}_z=\boldsymbol{L}_z^y$ along with their counterparts for $\boldsymbol{L}_z^x,\boldsymbol{L}_x^z$ and $\boldsymbol{L}_y^z$ by simple algebra.

Moreover, consider the SMW formula on $(\boldsymbol{Z}^T\mathbb{M}_y\boldsymbol{Z})^{-1}$, we have:

$$\begin{aligned}\boldsymbol{L}_z^y &= \boldsymbol{Z}(\boldsymbol{Z}^T(\boldsymbol{I}-\mathbb{P}_y)\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\\ &= \boldsymbol{Z}((\boldsymbol{Z}^T\boldsymbol{Z})^{-1}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{L}_y^z\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1})\boldsymbol{Z}^T\\ &= \mathbb{P}_z+\mathbb{P}_z\boldsymbol{L}_y^z\mathbb{P}_z.\end{aligned}$$

Equations for $\boldsymbol{L}_x^z, \boldsymbol{L}_z^x, \boldsymbol{L}_y^z$ by SMW formula can be calculated similarly.

Now we start our proof from observing that $\hat{\boldsymbol{Y}}_{res} = \mathbb{M}_z Y$ here (without $\boldsymbol{X}^T\boldsymbol{Z} = 0$). To show this, recall that $\hat{\boldsymbol{B}} = (\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}$ and $\hat{\boldsymbol{D}} = (\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Y}$, we have

$$\begin{aligned}\hat{\boldsymbol{Y}}_{res} &= \boldsymbol{Y}-\mathbb{P}_z\boldsymbol{X}\hat{\boldsymbol{B}}-\boldsymbol{Z}\hat{\boldsymbol{D}}\\ &= (\boldsymbol{I}-\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{M}_z-\boldsymbol{L}_z^x\mathbb{M}_x)\boldsymbol{Y}\\ &= (\boldsymbol{I}-\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{M}_z-(\mathbb{P}_z+\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{P}_z)\mathbb{M}_x)\boldsymbol{Y}\\ &= (\boldsymbol{I}-\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{M}_z-\mathbb{P}_z\mathbb{M}_x-\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{P}_z\mathbb{M}_x)\boldsymbol{Y}\\ &= (\boldsymbol{I}-\mathbb{P}_z\boldsymbol{L}_x^z\mathbb{M}_z-\mathbb{P}_z\mathbb{M}_x-\mathbb{P}_z(-\boldsymbol{L}_x^z\mathbb{M}_z+\mathbb{P}_x))\boldsymbol{Y}\\ &= (\boldsymbol{I}-\mathbb{P}_z\mathbb{M}_x-\mathbb{P}_z\mathbb{P}_x)\boldsymbol{Y}\\ &= (\boldsymbol{I}-\mathbb{P}_z)\boldsymbol{Y}\\ &= \mathbb{M}_z\boldsymbol{Y}. \qquad \text{(E.5)}\end{aligned}$$

Also, note that in the proof of (E.2), we have

$$\begin{aligned}&\boldsymbol{X}^T\mathbb{M}_y\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_y\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\ =&\boldsymbol{\zeta}-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \qquad \text{(E.6)}\end{aligned}$$

still holds true without the assumption $\boldsymbol{X}^T\boldsymbol{Z} = 0$. Combining (E.5) and (E.6) with (E.4), we can simplify our target to proof

$$\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}(-\hat{\boldsymbol{B}}^T\boldsymbol{\zeta}-\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')$$

$$=-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'$$

For the left-hand side,

$$\begin{aligned}
&\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}(-\hat{\boldsymbol{B}}^T\boldsymbol{\zeta}-\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{X}(\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}+\\
&\boldsymbol{Y}^T\mathbb{M}_x\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{L}_y^z\big(\mathbb{M}_z\boldsymbol{L}_x^z\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}+\mathbb{M}_x\boldsymbol{Z}(\boldsymbol{Z}^T\mathbb{M}_x\boldsymbol{Z})^{-1})\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{L}_y^z\big(\mathbb{M}_z\boldsymbol{L}_x^z\boldsymbol{Z}+\mathbb{M}_x\boldsymbol{Z}(\boldsymbol{I}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{L}_x^z\boldsymbol{Z})\big)(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{L}_y^z\big(\mathbb{M}_z\boldsymbol{L}_x^z+\mathbb{M}_x+\mathbb{M}_x\mathbb{P}_z\boldsymbol{L}_x^z\big)\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{L}_y^z\big(\mathbb{M}_z\boldsymbol{L}_x^z+\mathbb{M}_x+\mathbb{M}_x\boldsymbol{L}_x^z-\mathbb{M}_x\mathbb{M}_z\boldsymbol{L}_x^z\big)\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{L}_y^z\big(\mathbb{M}_z\boldsymbol{L}_x^z+\mathbb{M}_x+0+\mathbb{P}_x-\mathbb{M}_z\boldsymbol{L}_x^z\big)\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{Y}(\boldsymbol{Y}^T\mathbb{M}_z\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}',
\end{aligned}$$

match with the right-hand side, thus we finish the proof. □

F Proof of Lemma B.4

Proof.

$$
\begin{aligned}
&(\boldsymbol{I}+\boldsymbol{H}_{ir})^{-1}(\hat{\boldsymbol{\xi}}_{gls}-\boldsymbol{\zeta})+\boldsymbol{\zeta}\\
=&(\boldsymbol{I}+(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1})^{-1}((\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}'-\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')-\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}')+\\
&\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T+(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1})^{-1}(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}'-\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}')-\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}')+\\
&\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}'\\
=&\left((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}+\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\bar{\boldsymbol{y}}}'_{res}+\boldsymbol{\zeta}
\end{aligned}
$$

□

G Proof of Lemma B.5

Proof. By the SMW formula, we have

$$
\begin{aligned}
&((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}+\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\\
=&\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}-\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}}(\boldsymbol{S}+\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}})^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}.
\end{aligned}
$$

Since $(\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^T - \mathbb{P}_z\boldsymbol{X}\hat{\boldsymbol{B}} - \boldsymbol{Z}\hat{\boldsymbol{D}})^T(\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^T - \mathbb{P}_z\boldsymbol{X}\hat{\boldsymbol{B}} - \boldsymbol{Z}\hat{\boldsymbol{D}}) = \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}} + \boldsymbol{S}$, which gives

$$
\begin{aligned}
&\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T((\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1} + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1} \\
=&\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T(\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X} - \boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}}(\boldsymbol{S} + \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}})^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}) \\
=&\boldsymbol{S}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}}(\boldsymbol{S} + \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}})^{-1})\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X} \\
=&\boldsymbol{S}^{-1}(\boldsymbol{I} - (\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}} + \boldsymbol{S} - \boldsymbol{S})(\boldsymbol{S} + \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}})^{-1})\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X} \\
=&(\boldsymbol{S} + \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X}\hat{\boldsymbol{B}})^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X} \\
=&(\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res})^{-1}\hat{\boldsymbol{Y}}_{res}^T\boldsymbol{X}.
\end{aligned}
$$

Combining these results, we finished with

$$
\hat{\boldsymbol{\xi}}_{bay} = \boldsymbol{X}^T\hat{\boldsymbol{Y}}_{res}(\hat{\boldsymbol{Y}}_{res}^T\hat{\boldsymbol{Y}}_{res})^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}' - \hat{\boldsymbol{D}}^T\boldsymbol{Z}') + \boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}'.
$$

□

H Proof of Theorem 1

Proof. Theorem 1 (i) follows immediately from Lemmas B.3 to B.5. And Theorem 1 (ii) is already established by Lemmas B.2 and B.4. □

I Proof of Theorem 2

Proof. Given Assumption 1, when $n \to \infty$, we have the following results:

$$\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{Y}} = \boldsymbol{\alpha} + \bar{\boldsymbol{E}} \xrightarrow{p} \boldsymbol{\alpha}$$

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}^T \mathbb{M}_z \boldsymbol{X})^{-1} \boldsymbol{X}^T \mathbb{M}_z \boldsymbol{Y} = \boldsymbol{B} + (\boldsymbol{X}^T \mathbb{M}_z \boldsymbol{X})^{-1} \boldsymbol{X}^T \mathbb{M}_z \boldsymbol{E} \xrightarrow{p} \boldsymbol{B}$$

$$\hat{\boldsymbol{D}} = (\boldsymbol{Z}^T \mathbb{M}_x \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \mathbb{M}_x \boldsymbol{Y} = \boldsymbol{D} + (\boldsymbol{Z}^T \mathbb{M}_x \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \mathbb{M}_x \boldsymbol{E} \xrightarrow{p} \boldsymbol{D}$$

$$\eta \boldsymbol{S}^{-1} \sim W^{-1}(\eta, \eta \boldsymbol{\Gamma}^{-1}) \xrightarrow{p} \boldsymbol{\Gamma}^{-1},$$

where $\eta = n - p - p' - 1$. Denote $\boldsymbol{e}_j$ follows the distribution $F(0, \boldsymbol{\Gamma})$, then $\bar{\boldsymbol{y}}' \sim F\left(\boldsymbol{\alpha} + \boldsymbol{B}^T \boldsymbol{\xi} + \boldsymbol{D}^T \bar{\boldsymbol{z}}', l^{-1}\boldsymbol{\Gamma}\right)$. By the Slutsky's theorem, we have that

$$\lim_{n\to\infty} \hat{\boldsymbol{\xi}}_{gls} = \lim_{n\to\infty} (\hat{\boldsymbol{B}} \boldsymbol{S}^{-1} \hat{\boldsymbol{B}}^T)^{-1} \hat{\boldsymbol{B}} \boldsymbol{S}^{-1} (\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T \bar{\boldsymbol{z}}') \xrightarrow{d} F\left(\boldsymbol{\xi}, \frac{1}{l}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}\right). \tag{I.1}$$

Then, the asymptotic mean and variance of $\hat{\boldsymbol{\xi}}_{gls}$ are

$$\lim_{n\to\infty} \mathbb{E}(\hat{\boldsymbol{\xi}}_{gls}) = \boldsymbol{\xi} \quad \text{and} \quad \lim_{n\to\infty} \mathrm{Var}(\hat{\boldsymbol{\xi}}_{gls}) = l^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}.$$

Since we have assumed $\mathbb{E}||\boldsymbol{e}_i||_2^{2+\delta} = \mathbb{E}||\boldsymbol{e}_j||_2^{2+\delta} < \infty$ for a $\delta > 0$ by Assumption 1. (i), the uniform integrability of $\mathbb{E}||\boldsymbol{e}||_2^2$ is satisfied, which leads to the moment convergence as given above.

Furthermore, as $n \to \infty$,

$$\begin{aligned}
\mathbb{E}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) =& ((\boldsymbol{I} + \boldsymbol{H})^{-1} \mathbb{E}\hat{\boldsymbol{\xi}}_{gls} - \boldsymbol{\xi} - \{(\boldsymbol{I} + \boldsymbol{H})^{-1} - \boldsymbol{I}\}\boldsymbol{\zeta}) \\
\to& \{(\boldsymbol{I} + \boldsymbol{H})^{-1} - \boldsymbol{I}\}\boldsymbol{\xi} - \{(\boldsymbol{I} + \boldsymbol{H})^{-1} - \boldsymbol{I}\}\boldsymbol{\zeta}) \\
=& \{(\boldsymbol{I} + \boldsymbol{H})^{-1} - \boldsymbol{I}\}(\boldsymbol{\xi} - \boldsymbol{\zeta}), \\
\mathrm{var}(\hat{\boldsymbol{\xi}}) =& (\boldsymbol{I} + \boldsymbol{H})^{-1} \mathrm{var}(\hat{\boldsymbol{\xi}}_{gls})(\boldsymbol{I} + \boldsymbol{H}^T)^{-1} \\
\to& l^{-1}(\boldsymbol{I} + \boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I} + \boldsymbol{H}^T)^{-1},
\end{aligned}$$

and we finished the proof. □

J Proof of Theorem 3

Proof. From Theorem 2, the asymptotic bias and variance are

$$\text{Bias}(\hat{\boldsymbol{\xi}}, \boldsymbol{\xi}) \to ((\boldsymbol{I}+\boldsymbol{H})^{-1} - \boldsymbol{I})(\boldsymbol{\xi} - \boldsymbol{\zeta}) \quad \text{and} \quad \text{Var}(\hat{\boldsymbol{\xi}}) \to \frac{1}{l}(\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H})^{-1^T},$$

respectively. To derive the shrinkage estimator, we optimize the limiting average MSE with respect to $\boldsymbol{H}$,

$$\boldsymbol{H}_{opt} = \arg\min_{\boldsymbol{H}} \text{tr}\left(((\boldsymbol{I}+\boldsymbol{H})^{-1} - \boldsymbol{I})\mathbb{E}[(\boldsymbol{\xi}-\boldsymbol{\zeta})(\boldsymbol{\xi}-\boldsymbol{\zeta})^T]((\boldsymbol{I}+\boldsymbol{H})^{-1} - \boldsymbol{I})^T\right) + \frac{1}{l}\text{tr}\left((\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H})^{-1^T}\right).$$

The first order condition yield

$$\frac{\partial \text{MSE}(\hat{\boldsymbol{\xi}})}{\partial \boldsymbol{H}} : 2((\boldsymbol{I}+\boldsymbol{H})^{-1} - \boldsymbol{I})\mathbb{E}[(\boldsymbol{\xi}-\boldsymbol{\zeta})(\boldsymbol{\xi}-\boldsymbol{\zeta})^T] + \frac{2}{l}(\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1} = 0,$$

hence

$$(\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1} = \mathbb{E}[(\boldsymbol{\xi}-\boldsymbol{\zeta})(\boldsymbol{\xi}-\boldsymbol{\zeta})^T]\left\{(l)^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1} + \mathbb{E}[(\boldsymbol{\xi}-\boldsymbol{\zeta})(\boldsymbol{\xi}-\boldsymbol{\zeta})^T]\right\}^{-1},$$

which leads to $\tilde{\boldsymbol{H}}_{opt} = l^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}\left\{\boldsymbol{\Gamma}_x - \boldsymbol{\Gamma}_{xz}\boldsymbol{\Gamma}_z^{-1}\boldsymbol{\Gamma}_{xz}^T\right\}^{-1}$. □

K Proof of Theorem 4

To appreciate the benefit of the additional covariates $\boldsymbol{Z}$, suppose we omit $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ from the Models (2.3) and (2.4), and suppose that the remaining models are still correctly specified, say, $\boldsymbol{X}^T\boldsymbol{Z} = 0$, which prevents the estimation of $\boldsymbol{B}$ will suffer from the omitted variable bias. Then the proof can be given as follows.

Proof. First, for Theorem 4 (i)., we noticed that $\boldsymbol{\zeta} = 0$ and $\mathbb{M}z\boldsymbol{X} = \boldsymbol{X}$ given $\boldsymbol{X}^T\boldsymbol{Z} = 0$. By definition, when $n \to \infty$,

$$\begin{aligned}\mathbb{E}_\xi(\mathrm{MSE}(\hat{\boldsymbol{\xi}})) \to & \mathrm{tr}\left(((\boldsymbol{I}+\boldsymbol{H})^{-1}-\boldsymbol{I})\mathbb{E}_\xi(\boldsymbol{\xi}\boldsymbol{\xi}^T)((\boldsymbol{I}+\boldsymbol{H})^{-1}-\boldsymbol{I})^T\right)+\\ & l^{-1}\mathrm{tr}\left((\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H})^{-1^T}\right)\\ = & \mathrm{tr}\left(((\boldsymbol{I}+\boldsymbol{H})^{-1}-\boldsymbol{I})\boldsymbol{\Gamma}_x((\boldsymbol{I}+\boldsymbol{H})^{-1}-\boldsymbol{I})^T\right)+\\ & l^{-1}\mathrm{tr}\left((\boldsymbol{I}+\boldsymbol{H})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H})^{-1^T}\right).\end{aligned}$$

By a similar derivation of Theorem 3,

$$\tilde{\boldsymbol{H}}_{opt} = l^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}\boldsymbol{\Gamma}_x^{-1}.$$

When $n \to \infty$, we have $\boldsymbol{X}^T\boldsymbol{X}/n \to \boldsymbol{\Gamma}_x$. By plug-in $\boldsymbol{X}^T\boldsymbol{X}/n$, we noticed that $\boldsymbol{H}_{opt} \to \tilde{\boldsymbol{H}}_{opt}$, which means that $\hat{\boldsymbol{\xi}}_{opt}$ minimizes limiting $\mathbb{E}_\xi(\mathrm{MSE}(\hat{\boldsymbol{\xi}}))$ among all the $\hat{\boldsymbol{\xi}} = (\boldsymbol{I}+\boldsymbol{H})^{-1}(\hat{\boldsymbol{\xi}}_{gls}-\boldsymbol{\zeta})+\boldsymbol{\zeta}$.

Then, consider Theorem 4 (ii)., we will show that the limiting $\mathbb{E}(\mathrm{MSE}(\hat{\boldsymbol{\xi}}_{opt}))$ is positively correlated with the covariance term $\boldsymbol{\Gamma}$, and the covariance will be reduced with additional $\boldsymbol{Z}$ involved in.

The limiting MSE of $\hat{\boldsymbol{\xi}}_{opt}$ can be derived as

$$\begin{aligned}\mathbb{E}(\mathrm{MSE}(\hat{\boldsymbol{\xi}}_{opt})) \to & \mathrm{tr}(((\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1}-\boldsymbol{I})(n^{-1}\boldsymbol{X}^T\boldsymbol{X})((\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1}-\boldsymbol{I})^T)+\\ & \mathrm{tr}(l^{-1}(\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}(\boldsymbol{I}+\boldsymbol{H}_{opt}^T)^{-1})\\ = & \mathrm{tr}(((\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1})(n^{-1}\boldsymbol{X}^T\boldsymbol{X}+l^{-1}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1})(\boldsymbol{I}+\boldsymbol{H}_{opt}^T)^{-1})-\\ & 2\mathrm{tr}(((\boldsymbol{I}+\boldsymbol{H}_{opt})^{-1})(n^{-1}\boldsymbol{X}^T\boldsymbol{X}))+\mathrm{tr}(n^{-1}\boldsymbol{X}^T\boldsymbol{X})\\ = & \mathrm{tr}(n^{-1}\boldsymbol{X}^T\boldsymbol{X})-\\ & \mathrm{tr}\{((n^{-1}\boldsymbol{X}^T\boldsymbol{X})^{-1}+l^{-1}(n^{-1}\boldsymbol{X}^T\boldsymbol{X})^{-1}(\hat{\boldsymbol{B}}\boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(n^{-1}\boldsymbol{X}^T\boldsymbol{X})^{-1})^{-1}\},\end{aligned}$$

which is positively correlated with $\boldsymbol{\Gamma}$.

Given the full model

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\alpha}^T + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{Z}\boldsymbol{D} + \boldsymbol{E},$$

and a partial model

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\alpha}^T + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}_z,$$

with $\boldsymbol{E}_z = \boldsymbol{Z}\boldsymbol{D} + \boldsymbol{E}$. Since $\boldsymbol{X}^T\boldsymbol{Z} = 0$, the condition $\boldsymbol{X}^T\boldsymbol{E}_z = 0$ is also satisfied for the partial model, thus it is also correctly specified. However, the covariance term of $\boldsymbol{E}$ is $\boldsymbol{\Gamma}$, while for $\boldsymbol{E}_z$ is $\boldsymbol{\Gamma}_z = \boldsymbol{D}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{D} + \boldsymbol{\Gamma}$. Clearly $\boldsymbol{\Gamma}_z - \boldsymbol{\Gamma}$ is positive definite, and the limiting $\mathbb{E}(\text{MSE}(\hat{\boldsymbol{\xi}}_{opt}))$ is thus larger compared with the full model. □

L Proof of Theorem 5

Proof. Indeed, consider a class of penalized estimators solving

$$\tilde{\boldsymbol{\xi}} = \arg\min_{\boldsymbol{\xi}} \text{tr}[(\boldsymbol{Y}' - \hat{\boldsymbol{Y}}'(\boldsymbol{\xi}))\boldsymbol{S}^{-1}(\boldsymbol{Y}' - \hat{\boldsymbol{Y}}'(\boldsymbol{\xi}))^T] + \lambda\sigma^2(\boldsymbol{\xi}), \tag{L.1}$$

where $\sigma^2(\boldsymbol{\xi}) = 1/l + 1/n + \boldsymbol{\xi}^T\boldsymbol{C}_1\boldsymbol{\xi} - \boldsymbol{c}_2^T\boldsymbol{\xi} + c_3$, $\hat{\boldsymbol{Y}}'(\boldsymbol{\xi}) = \boldsymbol{1}_l\hat{\boldsymbol{\alpha}}^T + \boldsymbol{1}_l\boldsymbol{\xi}^T\hat{\boldsymbol{B}} + \boldsymbol{Z}'\hat{\boldsymbol{D}}$ and $\lambda \in \mathbb{R}$ is a tuning parameter controls the penalty. The first order condition yield

$$2\lambda\boldsymbol{C}_1\boldsymbol{\xi} - 2\lambda(\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1}\boldsymbol{\zeta} - 2l\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}' - \hat{\boldsymbol{B}}^T\boldsymbol{\xi}) = 0,$$

and it has the solution of the form

$$\tilde{\boldsymbol{\xi}} = \left(\frac{\lambda}{l}(\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1} + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\bar{\boldsymbol{y}}}'_{res} + \boldsymbol{\zeta}, \tag{L.2}$$

where $\hat{\bar{\boldsymbol{y}}}'_{res} = \bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'$.

When $\lambda = 0$,

$$\begin{aligned}\text{(L.2)} &= \left(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') + \boldsymbol{\zeta} \\ &= \left(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}'),\end{aligned}$$

coincide with (2.6).

When $\lambda = 1$,

$$
\begin{aligned}
\text{(L.2)} =& \left(\frac{1}{l}(\boldsymbol{X}^T\mathbb{M}_z\boldsymbol{X})^{-1} + \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\right)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{\zeta} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') + \boldsymbol{\zeta} \\
=&(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T + \frac{1}{l}(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1})^{-1}(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - \hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}')+ \\
&\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\
=&(\boldsymbol{I} + \frac{1}{l}(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}(\boldsymbol{X}^T\mathbb{M}_Z\boldsymbol{X})^{-1})^{-1}((\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}' - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{D}}^T\bar{\boldsymbol{z}}') - \\
&\boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}') + \boldsymbol{X}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\bar{\boldsymbol{z}}' \\
=&(\boldsymbol{I} + \boldsymbol{H}_{opt})^{-1}(\hat{\boldsymbol{\xi}}_{gls} - \boldsymbol{\zeta}) + \boldsymbol{\zeta},
\end{aligned}
$$

is the same as (2.11).

When $\lambda = l$, by Lemma B.4, we know $\hat{\boldsymbol{\xi}}_{ir}$ is the solution with $\lambda = l$. □

M Derivation of the optimal shrinkage estimator under heterogeneity

For the GLS estimator (2.16), we can derive the variance of $\hat{\boldsymbol{\xi}}_{gls}$ by

$$\text{var}(\hat{\boldsymbol{\xi}}_{gls}) = (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}.$$

If we further assume that $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}, \hat{\boldsymbol{\Gamma}}_1, \hat{\boldsymbol{\Gamma}}_2$ convergence to the true parameter, since $\mathbb{E}(\hat{\boldsymbol{A}}) = 0$, we can obtain the mean of $E(\hat{\boldsymbol{\xi}}_{gls}) = \boldsymbol{\xi}$. Combining the mean and variance term, we can yield

$$\hat{\boldsymbol{\xi}}_{gls} \sim \mathcal{N}(\boldsymbol{\xi}, (\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}).$$

Based on the asymptotic normal property, we then consider $\hat{\boldsymbol{\xi}} = \boldsymbol{H}\hat{\boldsymbol{\xi}}_{gls}$ as a shrinkage estimator of $\boldsymbol{\xi}$, where $\boldsymbol{H}$ is a $p \times p$ matrix. We have

$$\text{Bias}(\hat{\boldsymbol{\xi}}, \boldsymbol{\xi}) = (\boldsymbol{H} - \boldsymbol{I})\boldsymbol{\xi},$$

$$\text{Var}(\hat{\boldsymbol{\xi}}) = \boldsymbol{H}(\hat{\boldsymbol{B}}\boldsymbol{S}^{-1}\hat{\boldsymbol{B}}^T)^{-1}\boldsymbol{H}^T.$$

From the definition of limiting MSE, as $n \to \infty$,

$$\text{MSE}(\hat{\boldsymbol{\xi}}) \to \text{tr}\left((\boldsymbol{H} - \boldsymbol{I})\boldsymbol{\xi}\boldsymbol{\xi}^T(\boldsymbol{H} - \boldsymbol{I})^T\right) + \text{tr}\left(\boldsymbol{H}(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1}\boldsymbol{H}^T\right),$$

and $\hat{\boldsymbol{H}}_{opt} = \boldsymbol{\xi}\boldsymbol{\xi}^T\left\{(\boldsymbol{B}\boldsymbol{\Gamma}^{-1}\boldsymbol{B}^T)^{-1} + \boldsymbol{\xi}\boldsymbol{\xi}^T\right\}^{-1}$ that minimizes the limiting MSE. Then, replacing $\boldsymbol{\xi}\boldsymbol{\xi}^T$ by $\boldsymbol{X}^T\boldsymbol{X}/n$, we finally get the optimal shrinkage estimator

$$\hat{\boldsymbol{\xi}}_{opt} = \frac{\boldsymbol{X}^T\boldsymbol{X}}{n}\left((\hat{\boldsymbol{B}}\boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{B}}^T)^{-1} + \frac{\boldsymbol{X}^T\boldsymbol{X}}{n}\right)^{-1}\hat{\boldsymbol{\xi}}_{gls}, \quad \text{(M.1)}$$

which leads to (2.17).

N Simulation results for the convergence of the EM algorithm

We have conducted extra simulations to evaluate the performance of the parameter estimations in the linear effect models (2.15). The results are included here in Figure S1.

We have evaluated $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2$ under different sample sizes. The relative efficiency, defined as the ratio against the case when $n = 30$, is reported in Figure S1. When $n = 30$, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\Gamma}}_1$ were seen less stable, so there was a drastic drop in l_2 loss from $n = 30$ to $n = 100$. The l_2 losses of all parameter estimations decayed geometrically as n increases from 100. These observations agree with the existing theoretical investigations, e.g., those from Balakrishnan et al. (2017).

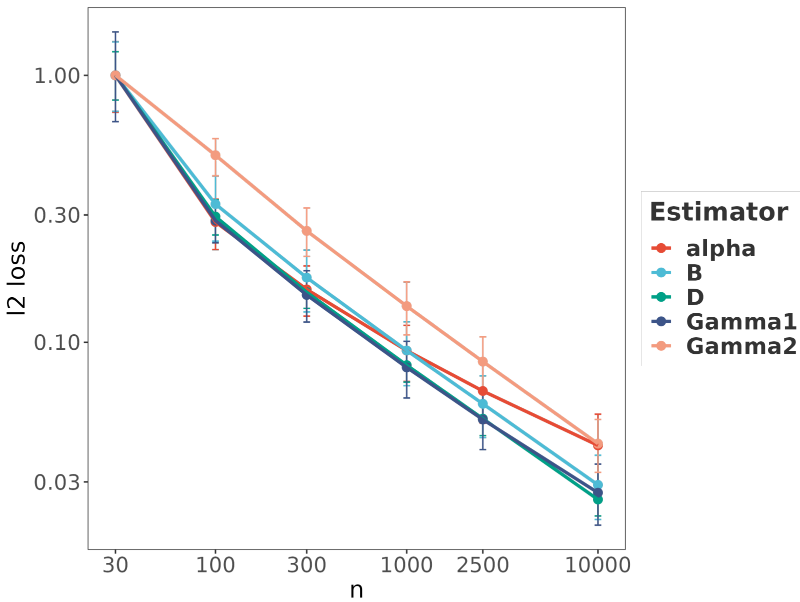

Figure S1: Empirical means and the 10th, 90th quantiles of the l_2 loss $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2$ (in log-scale) calculated from 1000 simulation replications for the estimators with respect to the sample size n.

O Simulation results against the σ_z

In section 3.1, we have introduced an extra hyperparameter σ_z to generate the repeated measurements. As shown in Figure S2, for the three statistical methods GLS, IR and OPT, the MSE was not affected by the σ_z, while the performance of RF was worse given a larger σ_z. This is expected from our Theorem 2 where the asymptotic mean and variance do not involve with the σ_z term.

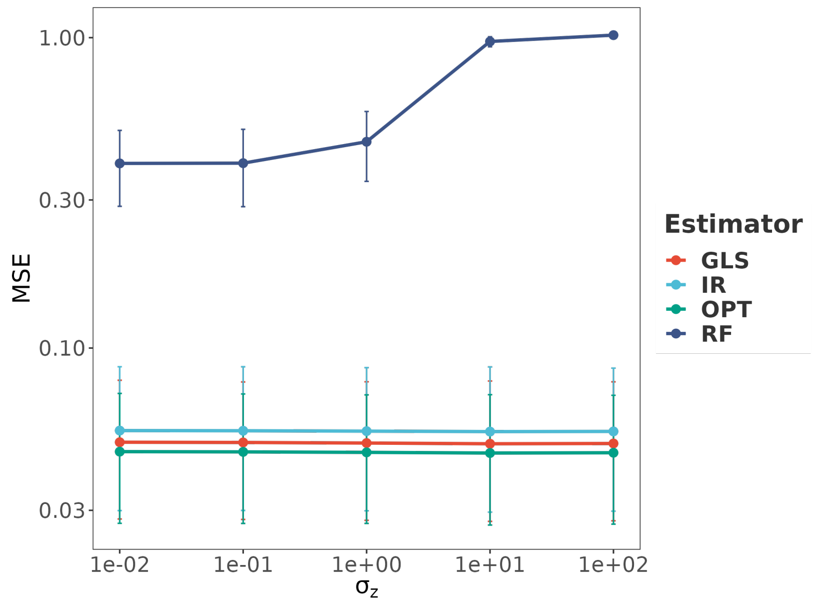

Figure S2: Empirical means and the 10th, 90th quantiles of the MSEs (in log-scale) calculated from 1000 simulation replications for the GLS, the inverse regression (IR), the optimal shrinkage (OPT) and the inverse regression with the random forest (RF) estimators with respect to the variances of the noises σ_z^2.

References

Balakrishnan, S., M. J. Wainwright, and B. Yu (2017). Statistical guarantees for the em algorithm: From population to sample-based analysis. *The Annals of Statistics 45*(1), 77–120, 44.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 1–22.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian Data*

Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Greene, W. (2008). *Econometric Analysis*. Pearson/Prentice Hall.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*(4), 963–974.

Meng, X.-L. and D. Van Dyk (1998). Fast em-type implementations for mixed effects models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 60*(3), 559–578.