ENVELOPE MODELS FOR PARSIMONIOIUS AND EFFICIENT MULTIVARIATE LINEAR REGRESSION

R. Dennis Cook[1], Bing Li[2] and Francesca Chiaromonte[2]

[1]*University of Minnesota and* [2]*Pennsylvania State University*

Supplementary Material

The referred equations in this Supplement are labeled as (S1.1), (S2.1) and so on, whereas labels such as (1), (2), etc. refer to equations the main text. We also include here additional lemmas and corollaries (and when necessary, their proofs), sometimes within the proof of an assertion in the main text. The proof of each assertion in the main text ends with a ■, and the proof of each additional lemma or corollary introduced in the Supplement ends with a □. Proofs are organized according to the sections in which the corresponding assertions appear.

S1 Proofs for Section 2

To prove the results in this section we need the following additional lemmas.

Lemma S1.1 *Let $\mathcal{R}$ be a u dimensional subspace of $\mathbb{R}^r$, and let $\mathbf{M} \in \mathbb{R}^{r\times r}$. $\mathcal{R}$ is an invariant subspace of $\mathbf{M}$ if and only if, for any $\mathbf{A} \in \mathbb{R}^{r\times s}$, $s \geq u$, such that $\mathrm{span}(\mathbf{A}) = \mathcal{R}$, there exists a $\mathbf{B} \in \mathbb{R}^{s\times s}$ such that $\mathbf{MA} = \mathbf{AB}$.*

Proof. Suppose there is a $\mathbf{B}$ that satisfies $\mathbf{MA} = \mathbf{AB}$. For every $\mathbf{v} \in \mathcal{R}$ there is a $\mathbf{t} \in \mathbb{R}^u$ so that $\mathbf{v} = \mathbf{At}$. Consequently, $\mathbf{Mv} = \mathbf{MAt} = \mathbf{ABt} \in \mathcal{R}$, which implies that $\mathcal{R}$ is an invariant subspace of $\mathbf{M}$.

Suppose that $\mathcal{R}$ is an invariant subspace of $\mathbf{M}$, and let $\mathbf{a}_j$, $j = 1, \ldots, s$ denote the columns of $\mathbf{A}$. Then $\mathbf{Ma}_j \in \mathcal{R}$, $j = 1, \ldots, s$. Consequently, $\mathrm{span}(\mathbf{MA}) \subseteq \mathcal{R}$, which implies there is a $\mathbf{B} \in \mathbb{R}^{s\times s}$ such that $\mathbf{MA} = \mathbf{AB}$. □

Lemma S1.2 *Let $\mathcal{R}$ reduce $\mathbf{M} \in \mathbb{R}^{r\times r}$. Then $\mathbf{M}\mathcal{R} = \mathcal{R}$ if and only if $\mathcal{R} \subseteq \mathrm{span}(\mathbf{M})$.*

Proof. Assume that $\mathbf{M}\mathcal{R} = \mathcal{R}$. Then, with $\mathbf{A}$ as defined in Corollary 2.1, $\mathbf{MA} = \mathbf{AB}$ for some full rank matrix $\mathbf{B} \in \mathbb{R}^{u\times u}$. Consequently, $\mathbf{A}^T\mathbf{MA}$ is full rank. It follows from Corollary 2.1 that $\mathcal{R} \subseteq \mathrm{span}(\mathbf{M})$.

Assume that $\mathcal{R} \subseteq \mathrm{span}(\mathbf{M})$. Then it follows from Corollary 2.1 that $\mathbf{A}^T\mathbf{MA}$ is of full rank. Thus, $\mathbf{B}$ must have full rank in the representation $\mathbf{MA} = \mathbf{AB}$, which implies $\mathbf{M}\mathcal{R} = \mathcal{R}$. □

Lemma S1.3 *Suppose that $\mathcal{R}$ reduces $\mathbf{M} \in \mathbb{S}^{r\times r}$. Then $\mathbf{M}$ has a spectral decomposition with eigenvectors in $\mathcal{R}$ or in $\mathcal{R}^\perp$.*

PROOF. Let $\mathbf{A}_0 \in \mathbb{R}^{r\times u}$ be a semi-orthogonal matrix whose columns span $\mathcal{R}$ and let $\mathbf{A}_1$ be its completion, such that $(\mathbf{A}_0, \mathbf{A}_1) \equiv \mathbf{A}$ is an orthogonal matrix. Because $\mathbf{M}\mathcal{R} \subseteq \mathcal{R}$ and $\mathbf{M}\mathcal{R}^\perp \subseteq \mathcal{R}^\perp$, it follows from Lemma S1.1 there exist matrices $\mathbf{B}_0 \in \mathbb{R}^{u\times u}$ and $\mathbf{B}_1 \in \mathbb{R}^{(r-u)\times(r-u)}$ such that $\mathbf{M}\mathbf{A}_0 = \mathbf{A}_0\mathbf{B}_0$ and $\mathbf{M}\mathbf{A}_1 = \mathbf{A}_1\mathbf{B}_1$. Hence

$$\mathbf{M}\begin{pmatrix}\mathbf{A}_0 & \mathbf{A}_1\end{pmatrix} = \begin{pmatrix}\mathbf{A}_0 & \mathbf{A}_1\end{pmatrix}\begin{pmatrix}\mathbf{B}_0 & 0\\ 0 & \mathbf{B}_1\end{pmatrix} \Leftrightarrow \mathbf{M} = \mathbf{A}\begin{pmatrix}\mathbf{B}_0 & 0\\ 0 & \mathbf{B}_1\end{pmatrix}\mathbf{A}^T.$$

Because $\mathbf{M}$ is symmetric, so must be $\mathbf{B}_0$ and $\mathbf{B}_1$. Hence $\mathbf{B}_0$ and $\mathbf{B}_1$ have spectral decompositions $\mathbf{C}_0\mathbf{\Lambda}_0\mathbf{C}_0^T$ and $\mathbf{C}_1\mathbf{\Lambda}_1\mathbf{C}_1^T$ for some diagonal matrices $\mathbf{\Lambda}_0$ and $\mathbf{\Lambda}_1$ and orthogonal matrices $\mathbf{C}_0$ and $\mathbf{C}_1$. Let $\mathbf{C} = \operatorname{diag}(\mathbf{C}_0, \mathbf{C}_1)$ and $\mathbf{\Lambda} = \operatorname{diag}(\mathbf{\Lambda}_0, \mathbf{\Lambda}_1)$. Then,

$$\mathbf{M} = \mathbf{A}\mathbf{C}\mathbf{\Lambda}\mathbf{C}^T\mathbf{A}^T \equiv \mathbf{D}\mathbf{\Lambda}\mathbf{D}^T, \tag{S1.1}$$

where $\mathbf{D} = \mathbf{A}\mathbf{C}$. The first u columns of $\mathbf{D}$, which form the matrix $\mathbf{A}_0\mathbf{C}_0$, span $\mathcal{R}$. Moreover, $\mathbf{D}$ is an orthogonal matrix, and thus (S1.1) is a spectral decomposition of $\mathbf{M}$ with eigenvectors in $\mathcal{R}$ or $\mathcal{R}^\perp$. □

Proof of Proposition 2.1 Assume that $\mathbf{M}$ can be written as in (4). Then for any $\mathbf{v} \in \mathcal{R}$, $\mathbf{M}\mathbf{v} \in \mathcal{R}$, and for and $\mathbf{v} \in \mathcal{R}^\perp$, $\mathbf{M}\mathbf{v} \in \mathcal{R}^\perp$. Consequently, $\mathcal{R}$ reduces $\mathbf{M}$.

Next, assume that $\mathcal{R}$ reduces $\mathbf{M}$. We must show that $\mathbf{M}$ satisfies (4). Let $u = \dim(\mathcal{R})$. It follows from Lemma S1.1 that there is a $\mathbf{B} \in \mathbb{R}^{u\times u}$ that satisfies $\mathbf{M}\mathbf{A} = \mathbf{A}\mathbf{B}$, where $\mathbf{A} \in \mathbb{R}^{r\times u}$ and $\operatorname{span}(\mathbf{A}) = \mathcal{R}$. This implies $\mathbf{Q}_\mathcal{R}\mathbf{M}\mathbf{A} = 0$ which is equivalent to $\mathbf{Q}_\mathcal{R}\mathbf{M}\mathbf{P}_\mathcal{R} = 0$. By the same logic applied to $\mathcal{R}^\perp$, $\mathbf{P}_\mathcal{R}\mathbf{M}\mathbf{Q}_\mathcal{R} = 0$. Consequently,

$$\mathbf{M} = (\mathbf{P}_\mathcal{R} + \mathbf{Q}_\mathcal{R})\mathbf{M}(\mathbf{P}_\mathcal{R} + \mathbf{Q}_\mathcal{R}) = \mathbf{P}_\mathcal{R}\mathbf{M}\mathbf{P}_\mathcal{R} + \mathbf{Q}_\mathcal{R}\mathbf{M}\mathbf{Q}_\mathcal{R}. \qquad \blacksquare$$

Proof of Corollary 2.1 The first conclusion follows immediately from Proposition 2.1.

To show the second conclusion, first assume that $\mathbf{A}^T\mathbf{M}\mathbf{A}$ is full rank. Then, from Lemma S1.1, $\mathbf{B}$ must be full rank in the representation $\mathbf{M}\mathbf{A} = \mathbf{A}\mathbf{B}$. Consequently, any vector in $\mathcal{R}$ can be written as a linear combination of the columns of $\mathbf{M}$ and thus $\mathcal{R} \subseteq \operatorname{span}(\mathbf{M})$. Next, assume that $\mathcal{R} \subseteq \operatorname{span}(\mathbf{M})$. Then there is a full rank matrix $\mathbf{V} \in \mathbb{R}^{r\times u}$ such that $\mathbf{M}\mathbf{V} = \mathbf{A}$ and thus that $\mathbf{A}^T\mathbf{M}\mathbf{V} = \mathbf{I}_u$. Substituting $\mathbf{M}$ from Proposition 2.1, we have $(\mathbf{A}^T\mathbf{M}\mathbf{A})(\mathbf{A}^T\mathbf{V}) = \mathbf{I}_u$. It follows that $\mathbf{A}^T\mathbf{M}\mathbf{A}$ is of full rank.

For the third conclusion, since $\mathbf{M}$ is full rank $\mathcal{R} \subseteq \operatorname{span}(\mathbf{M})$ and $\mathcal{R}^\perp \subseteq \operatorname{span}(\mathbf{M})$. Consequently, both $\mathbf{A}^T\mathbf{M}\mathbf{A}$ and $\mathbf{A}_0^T\mathbf{M}\mathbf{A}_0$ are full rank. Thus the right hand side of (5) is defined. Meanwhile, note that $\mathbf{P}_\mathcal{R} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{Q}_\mathcal{R} = \mathbf{A}_0\mathbf{A}_0^T$. Hence, by (4), $\mathbf{M} = \mathbf{A}\mathbf{A}^T\mathbf{M}\mathbf{A}\mathbf{A}^T + \mathbf{A}_0\mathbf{A}_0^T\mathbf{M}\mathbf{A}_0\mathbf{A}_0^T$. Multiply this and the right hand side of (5) to complete the proof. ■

Proof of Proposition 2.2 The equivalence of 1 and 4 is known and can be found in Conway (1990, page 39). We now demonstrate the equivalence of 1, 2, and 3.

1 implies 2: If $\mathbf{v} \in \mathcal{R}$, then

$$\mathbf{v} = \mathbf{I}_p \mathbf{v} = \left(\sum_{i=1}^{q} \mathbf{P}_i \right) \mathbf{v} = \sum_{i=1}^{q} \mathbf{P}_i \mathbf{v} \in \oplus_{i=1}^{q} \mathbf{P}_i \mathcal{R}.$$

Hence $\mathcal{R} \subseteq \oplus_{i=1}^{q} \mathbf{P}_i \mathcal{R}$. Conversely, if $\mathbf{v} \in \oplus_{i=1}^{q} \mathbf{P}_i \mathcal{R}$, then $\mathbf{v}$ can be written as a linear combination of $\mathbf{P}_1 \mathbf{v}_1, \ldots, \mathbf{P}_q \mathbf{v}_q$ where $\mathbf{v}_1, \ldots, \mathbf{v}_q$ belong to $\mathcal{R}$. By Lemma S1.3, $\mathbf{P}_i \mathbf{w} \in \mathcal{R}$ for any $\mathbf{w} \in \mathcal{R}$. Hence any linear combination of $\mathbf{P}_1 \mathbf{v}_1, \ldots, \mathbf{P}_q \mathbf{v}_q$, with $\mathbf{v}_1, \ldots, \mathbf{v}_q$ belonging to $\mathcal{R}$, belongs to $\mathcal{R}$. That is, $\oplus_{i=1}^{q} \mathbf{P}_i \mathcal{R} \subseteq \mathcal{R}$.

2 implies 3: If $\mathbf{v} \in \mathcal{R}$ then, from the previous step, $\mathbf{P}_i \mathbf{v} \in \mathcal{R}$, $i = 1, \ldots, q$. Hence

$$\left(\sum_{i=1}^{q} \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i \right) \mathbf{v} = \sum_{i=1}^{q} \mathbf{P}_i \mathbf{v} = \mathbf{v} = \mathbf{P}_{\mathcal{R}} \mathbf{v}.$$

Now let $\mathbf{v} \in \mathcal{R}^{\perp}$. Then, $\mathbf{v} \perp \mathbf{P}_i \mathcal{R}$ for each i. Because $\mathbf{P}_i$ is self-adjoint we see that $\mathbf{P}_i \mathbf{v} \perp \mathcal{R}$ for each i. Consequently

$$\left(\sum_{i=1}^{q} \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i \right) \mathbf{v} = 0 = \mathbf{P}_{\mathcal{R}} \mathbf{v}.$$

It follows that $(\sum \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i)\mathbf{v} = \mathbf{P}_{\mathcal{R}} \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^r$. Hence the two matrices are the same.

3 implies 1: Again, if $\mathbf{v} \in \mathcal{R}$ then $\mathbf{P}_i \mathbf{v} \in \mathcal{R}$, $i = 1, \ldots, q$. Hence, indicating with m_i, $i = 1, \ldots, q$ the distinct eigenvalues of M we have

$$\mathbf{P}_{\mathcal{R}} \mathbf{M} \mathbf{v} = \sum_{i=1}^{q} m_i \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i \mathbf{v} = \sum_{i=1}^{q} m_i \mathbf{P}_i \mathbf{P}_{\mathcal{R}} \mathbf{P}_i \mathbf{v} = \mathbf{M} \mathbf{v}.$$

It follows that $\mathbf{M} \mathcal{R} \subseteq \mathcal{R}$. ■

Proof of Proposition 2.3 To prove that $\oplus_{i=1}^{q} \mathbf{P}_i \mathcal{S}$ is the smallest reducing subspace of $\mathbf{M}$ that contains $\mathcal{S}$, it suffices to prove the following statements:

1. $\oplus_{i=1}^{q} \mathbf{P}_i \mathcal{S}$ reduces $\mathbf{M}$.

2. $\mathcal{S} \subseteq \oplus_{i=1}^{q} \mathbf{P}_i \mathcal{S}$.

3. If $\mathcal{T}$ reduces $\mathbf{M}$ and $\mathcal{S} \subseteq \mathcal{T}$, then $\oplus_{i=1}^{q} \mathbf{P}_i \mathcal{S} \subseteq \mathcal{T}$.

Statement 1 follows from Proposition 2.2, as applied to $\mathcal{R} \equiv \oplus_{i=1}^q \mathbf{P}_i\mathcal{S}$. Statement 2 holds because $\mathcal{S} = \{\mathbf{P}_1\mathbf{v} + \cdots + \mathbf{P}_q\mathbf{v} : \mathbf{v} \in \mathcal{S}\} \subseteq \oplus_{i=1}^q \mathbf{P}_i\mathcal{S}$. Turning to statement 3, if $\mathcal{T}$ reduces $\mathbf{M}$, it can be written as $\mathcal{T} = \oplus_{i=1}^q \mathbf{P}_i\mathcal{T}$ by Proposition 2.2. If, in addition, $\mathcal{S} \subseteq \mathcal{T}$ then we have $\mathbf{P}_i\mathcal{S} \subseteq \mathbf{P}_i\mathcal{T}$ for $i = 1, \ldots, q$. Statement 3 follows since $\oplus_{i=1}^q \mathbf{P}_i\mathcal{S} \subseteq \oplus_{i=1}^q \mathbf{P}_i\mathcal{T} = \mathcal{T}$.

■

Proof of Proposition 2.4 Because $\mathbf{K}$ and $\mathbf{M}$ commute, they can be diagonalized simultaneously by an orthogonal matrix, say $\mathbf{U}$. Recall that $\mathbf{P}_i$ is the projection on the i-th eigenspace of $\mathbf{M}$, and let $d_i = \text{rank}(\mathbf{P}_i)$. Partition $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_q)$, where $\mathbf{U}_i$ contains d_i columns, $i = 1, \ldots, q$. Without loss of generality, we can assume that $\mathbf{U}_i\mathbf{U}_i^T = \mathbf{P}_i$ for $i = 1, \ldots, q$. Then $\mathbf{K}$ can be written as $\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^T + \cdots \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^T$, where the $\mathbf{\Lambda}_i$'s are diagonal matrices of dimension $d_i \times d_i$. It follows that

$$\begin{aligned}\mathbf{K}\mathbf{P}_i =& (\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^T + \cdots + \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^T)\mathbf{U}_i\mathbf{U}_i^T \\ =& \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^T = \mathbf{U}_i\mathbf{U}_i^T(\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^T + \cdots \mathbf{U}_q\mathbf{\Lambda}_q\mathbf{U}_q^T) = \mathbf{P}_i\mathbf{K}.\end{aligned}$$

That is, $\mathbf{K}$ and $\mathbf{P}_i$ commute. Now, by Proposition 2.2, $\mathcal{E}_{\mathbf{M}}(\mathcal{S}) = \oplus_i^q \mathbf{P}_i\mathcal{S}$. Hence

$$\begin{aligned}\mathbf{K}\mathcal{E}_{\mathbf{M}}(\mathcal{S}) =& \{\mathbf{K}\mathbf{P}_1\mathbf{h}_1 + \cdots + \mathbf{K}\mathbf{P}_q\mathbf{h}_q : \mathbf{h}_1, \ldots, \mathbf{h}_q \in \mathcal{S}\} \\ =& \{\mathbf{P}_1\mathbf{K}\mathbf{h}_1 + \cdots + \mathbf{P}_q\mathbf{K}\mathbf{h}_q : \mathbf{h}_1, \ldots, \mathbf{h}_q \in \mathcal{S}\} = \oplus_{i=1}^q \mathbf{P}_i\mathbf{K}\mathcal{S}.\end{aligned}$$

By Proposition 2.2 again, the right hand side is $\mathcal{E}_{\mathbf{M}}(\mathbf{K}\mathcal{S})$.

Now suppose, in addition, that $\mathcal{S} \subseteq \text{span}(\mathbf{K})$ and $\mathcal{S}$ reduces $\mathbf{K}$. We note that if $\mathbf{K}$ commutes with $\mathbf{M}$, then span($\mathbf{K}$) reduces $\mathbf{M}$. This is because, for all $\mathbf{h} \in \mathbb{R}^r$, $\mathbf{M}\mathbf{K}\mathbf{h} = \mathbf{K}\mathbf{M}\mathbf{h} \subseteq \text{span}(\mathbf{K})$. Hence $\mathcal{E}_{\mathbf{M}}(\mathcal{S}) \subseteq \text{span}(\mathbf{K})$. By Lemma S1.2, then, $\mathbf{K}\mathcal{E}_{\mathbf{M}}(\mathcal{S}) = \mathcal{E}_{\mathbf{M}}(\mathcal{S})$, which, in conjunction with (6), implies (7). ■

S2 Proofs for Section 3

Proof of Proposition 3.1 We need to show that $\mathbf{\Sigma}^{-1}\mathcal{B} = \mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B}$, and

$$\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}) = \mathcal{E}_{\mathbf{\Sigma}_{\mathbf{Y}}}(\mathcal{B}) = \mathcal{E}_{\mathbf{\Sigma}}(\mathbf{\Sigma}^{-1}\mathcal{B}) = \mathcal{E}_{\mathbf{\Sigma}_{\mathbf{Y}}}(\mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B}) = \mathcal{E}_{\mathbf{\Sigma}_{\mathbf{Y}}}(\mathbf{\Sigma}^{-1}\mathcal{B}) = \mathcal{E}_{\mathbf{\Sigma}}(\mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B}).$$

Model (1) implies model (10) by construction and thus $\mathbf{\Sigma}_{\mathbf{Y}} = \mathbf{\Sigma} + \mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^T$, where $\mathbf{V} = \boldsymbol{\eta}\text{var}(\mathbf{X})\boldsymbol{\eta}^T$. By matrix multiplication we can show that

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{Y}}^{-1} =& \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{\Gamma}(\mathbf{V}^{-1} + \mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma})\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}, \\ \mathbf{\Sigma}^{-1} =& \mathbf{\Sigma}_{\mathbf{Y}}^{-1} - \mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathbf{\Gamma}(-\mathbf{V}^{-1} + \mathbf{\Gamma}^T\mathbf{\Sigma}_{\mathbf{Y}}^{-1}\mathbf{\Gamma})\mathbf{\Gamma}^T\mathbf{\Sigma}_{\mathbf{Y}}^{-1}.\end{aligned}$$

The first equality implies $\mathrm{span}(\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{\Gamma}) \subseteq \mathrm{span}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma})$; the second implies $\mathrm{span}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}) \subseteq \mathrm{span}(\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{\Gamma})$. Hence $\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B} = \boldsymbol{\Sigma}^{-1}\mathcal{B}$, recalling that $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ by construction. From this we also deduce $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{Y}}}(\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B}) = \mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{Y}}}(\boldsymbol{\Sigma}^{-1}\mathcal{B})$ and $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\mathcal{B}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}^{-1}\mathcal{B})$.

We next show that $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}) = \mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{Y}}}(\mathcal{B})$ by demonstrating that $\mathcal{R} \subseteq \mathbb{R}^p$ is a reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathcal{B}$ if and only if it is a reducing subspace of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ that contains $\mathcal{B}$. Suppose $\mathcal{R}$ is a reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathcal{B}$. Let $\boldsymbol{\alpha} \in \mathcal{R}$. Then $\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\alpha} = \boldsymbol{\Sigma}\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathrm{var}(\mathbf{X})\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\boldsymbol{\alpha}$. $\boldsymbol{\Sigma}\boldsymbol{\alpha} \in \mathcal{R}$ because $\mathcal{R}$ reduces $\boldsymbol{\Sigma}$; the second term on the right is a vector in $\mathcal{R}$ because $\mathcal{B} \subseteq \mathcal{R}$. Thus, $\mathcal{R}$ is a reducing subspace of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ and by construction it contains $\mathcal{B}$. Next, suppose $\mathcal{R}$ is a reducing subspace of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ that contains $\mathcal{B}$. The reverse implication follows similarly by reasoning in terms of $\boldsymbol{\Sigma}\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\alpha} - \boldsymbol{\Gamma}\boldsymbol{\eta}\mathrm{var}(\mathbf{X})\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\boldsymbol{\alpha}$. We have $\boldsymbol{\Sigma}_{\mathbf{Y}}\boldsymbol{\alpha} \in \mathcal{R}$ because $\mathcal{R}$ reduces $\boldsymbol{\Sigma}_{\mathbf{Y}}$; the second term on the right is a vector in $\mathcal{R}$ because $\mathcal{B} \subseteq \mathcal{R}$. The remaining equalities follow immediately from (8). ■

S3 Proofs for Section 4

Proof of Lemma 4.1 From the properties of a projection we see that, for any $\mathbf{B} \in \mathcal{A}$, we have $\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}\mathbf{B}^T\mathbf{P}_{\mathbf{V}} = \mathbf{B}^T$. It follows that

$$\mathrm{tr}(\mathbf{A}^*\boldsymbol{\Lambda}\mathbf{B}^T) = \mathrm{tr}[\mathbf{P}_{\mathbf{V}}\mathbf{U}\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}^T\boldsymbol{\Lambda}\mathbf{B}^T] = \mathrm{tr}[\mathbf{U}\boldsymbol{\Lambda}\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}\mathbf{B}^T\mathbf{P}_{\mathbf{V}}] = \mathrm{tr}(\mathbf{U}\boldsymbol{\Lambda}\mathbf{B}^T).$$

Thus $\mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}\mathbf{B}] = 0$. Now decompose the objective function (11) as

$$\mathrm{tr}[(\mathbf{U} - \mathbf{A})\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A})^T] = \mathrm{tr}[(\mathbf{U} - \mathbf{A}^* + \mathbf{A}^* - \mathbf{A})\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A}^* + \mathbf{A}^* - \mathbf{A})^T]$$

Because $\mathbf{A}^* - \mathbf{A} \in \mathcal{A}$, the cross product term in the above is $\mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}(\mathbf{A}^* - \mathbf{A})^T] = 0$. Hence

$$\begin{aligned}\mathrm{tr}[(\mathbf{U} - \mathbf{A})\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A})^T] =& \mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A}^*)^T] + \mathrm{tr}[(\mathbf{A}^* - \mathbf{A})\boldsymbol{\Lambda}(\mathbf{A}^* - \mathbf{A})^T] \\ \geq& \mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A}^*)^T].\end{aligned}$$

The lower bound is achieved when $\mathbf{A} = \mathbf{A}^*$, in which case

$$\mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}(\mathbf{U} - \mathbf{A}^*)^T] = \mathrm{tr}[(\mathbf{U} - \mathbf{A}^*)\boldsymbol{\Lambda}\mathbf{U}^T] = \mathrm{tr}(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T) - \mathrm{tr}(\mathbf{A}^*\boldsymbol{\Lambda}\mathbf{U}^T).$$

However, by definition of $\mathbf{A}^*$ and the property of projection,

$$\mathrm{tr}(\mathbf{A}^*\boldsymbol{\Lambda}\mathbf{U}^T) = \mathrm{tr}(\mathbf{P}_{\mathbf{V}}\mathbf{U}\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}^T\boldsymbol{\Lambda}\mathbf{U}^T) = \mathrm{tr}[\mathbf{P}_{\mathbf{V}}\mathbf{U}\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}^T\boldsymbol{\Lambda}\mathbf{P}_{\mathbf{W}(\boldsymbol{\Lambda})}\mathbf{U}^T\mathbf{P}_{\mathbf{V}}],$$

as desired. ■

Proof of Lemma 4.1 Straightforward and omitted.

Proof of Lemma 4.3 Because $\mathbf{P}$ is the projection onto span($\mathbf{A}$), we have

$$L(\mathbf{A}) = [\det_0(\mathbf{A})]^{-\frac{1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{UPA}^{\dagger}\mathbf{PU}^T)}. \tag{S3.2}$$

If we write $\mathbf{U}^T = (\mathbf{U}_1, \ldots, \mathbf{U}_n)$, then the above is proportional to the likelihood of $\mathbf{U}_1, \ldots, \mathbf{U}_n$ if they are iid $N(0, \mathbf{A})$. The maximum likelihood estimator (among all $\mathbf{A}$) is the sample variance $\mathbf{A}^* = \mathbf{PU}^T\mathbf{UP}/n$. Note that $\mathbf{A}^*$ happens to be in $\mathcal{A}$. Therefore $L(\mathbf{A})$ is maximized by $\mathbf{A}^*$ among $\mathbf{A}$. In the meantime,

$$\mathrm{tr}[\mathbf{UP}(\mathbf{A}^*)^{\dagger}\mathbf{PU}^T] = \mathrm{tr}[(\mathbf{A}^*)^{\dagger}\mathbf{PU}^T\mathbf{UP}] = n\mathrm{tr}[(\mathbf{A}^*)^{\dagger}\mathbf{A}^*] = nk,$$

where the last equality holds because $(\mathbf{A}^*)^{\dagger}\mathbf{A}^*$ is a projection matrix of rank k. Substitute the above equality into (S3.2) and use the relation $n^k\det_0(\mathbf{A}^*) = \det_0(\mathbf{PU}^T\mathbf{UP})$ to complete the proof. ∎

S4 Proofs for Section 5

Proof of Theorem 5.1 The asymptotic distribution (23), where $\mathbf{H}$ is as defined in (22), follows from Shapiro (1986, Proposition 4.1). To prove the equality $\mathbf{V}_0 \leq \mathbf{V}$ we note that

$$\mathbf{V} - \mathbf{V}_0 = \mathbf{J}^{-1} - \mathbf{H}(\mathbf{H}^T\mathbf{JH})^{\dagger}\mathbf{H}^T = \mathbf{J}^{-\frac{1}{2}}[\mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\mathbf{H}(\mathbf{H}^T\mathbf{JH})^{\dagger}\mathbf{H}^T\mathbf{J}^{\frac{1}{2}}]\mathbf{J}^{-\frac{1}{2}}. \tag{S4.3}$$

Since the matrix $\mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\mathbf{H}(\mathbf{H}^T\mathbf{JH})^{\dagger}\mathbf{H}^T\mathbf{J}^{\frac{1}{2}}$ is the projection on to orthogonal complement of span($\mathbf{J}^{\frac{1}{2}}\mathbf{H}$) relative to the standard inner product, it is positive semidefinite, which implies that $\mathbf{V} - \mathbf{V}_0$ is positive semidefinite. From (S4.3) we can also see that

$$\mathbf{V}^{-\frac{1}{2}}(\mathbf{V} - \mathbf{V}_0)\mathbf{V}^{-\frac{1}{2}} = \mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\mathbf{H}(\mathbf{H}^T\mathbf{JH})^{\dagger}\mathbf{H}^T\mathbf{J}^{\frac{1}{2}} = \mathbf{Q}_{\mathbf{J}^{\frac{1}{2}}\mathbf{H}},$$

which proves the last statement of the theorem. We still need to derive the an explicit expression for $\mathbf{H}$ as given by (24). To do so we need to find expressions for the eight partial derivatives $\partial\mathbf{h}_i/\partial\boldsymbol{\phi}_j^T$, $i = 1, 2$, $j = 1, 2, 3, 4$. We divide these derivations into two steps.

Step 1: Compute $\partial\mathbf{h}_1/\partial\boldsymbol{\phi}^T$

First,

$$\frac{\partial\mathbf{h}_1}{\partial\boldsymbol{\phi}_1^T} = \frac{\partial\mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\eta})}{\partial\mathrm{vec}^T(\boldsymbol{\eta})} = \frac{\partial[(\mathbf{I}_p \otimes \boldsymbol{\Gamma})\mathrm{vec}(\boldsymbol{\eta})]}{\partial\mathrm{vec}^T(\boldsymbol{\eta})} = \mathbf{I}_p \otimes \boldsymbol{\Gamma} \in \mathbb{R}^{pr\times pu}.$$

In a similar way,

$$\frac{\partial\mathbf{h}_1}{\partial\boldsymbol{\phi}_2^T} = \frac{\partial\mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\eta})}{\partial\mathrm{vec}^T(\boldsymbol{\Gamma})} = \boldsymbol{\eta}^T \otimes \mathbf{I}_r \in \mathbb{R}^{pr\times ur}.$$

Clearly, $\partial \mathbf{h}_1/\partial \boldsymbol{\phi}_3^T = \mathbf{0}$, $\partial \mathbf{h}_1/\partial \boldsymbol{\phi}_4^T = \mathbf{0}$, where the first matrix has dimensions $pr \times u(u+1)/2$, and the second matrix has dimensions $pr \times (r-u)(r-u+1)/2$.

Step 2: Compute $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}^T$

Since $\mathbf{h}_2$ does not depend on $\boldsymbol{\phi}_1$ we have $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}_1^T = \mathbf{0}$. Note that this matrix is of dimension $r(r+1)/2 \times pu$.

To compute $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}_2^T$, let $\mathbf{h}_{21}(\boldsymbol{\phi}) = \mathrm{vech}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T)$ and $\mathbf{h}_{22}(\boldsymbol{\phi}) = \mathrm{vech}(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)$, so that we can write

$$\frac{\partial \mathbf{h}_2}{\partial \boldsymbol{\phi}_2^T} = \frac{\partial \mathbf{h}_{21}}{\partial \boldsymbol{\phi}_2^T} + \frac{\partial \mathbf{h}_{22}}{\partial \boldsymbol{\phi}_2^T}. \tag{S4.4}$$

where $\boldsymbol{\phi}_2 = \mathrm{vec}(\boldsymbol{\Gamma})$. The following two lemmas, which are presented without proof, will facilitate computation of the derivatives in (S4.4).

Lemma D.1 *Let $\mathbf{X}$ be a matrix of arbitrary dimensions, and let $\mathbf{F}(\mathbf{X}) \in \mathbb{R}^{m\times p}$ and $\mathbf{G}(\mathbf{X}) \in \mathbb{R}^{p\times q}$ be matrix-valued differentiable function of $\mathbf{X}$. Then*

$$\frac{\partial \mathrm{vec}[\mathbf{F}(\mathbf{X})\mathbf{G}(\mathbf{X})]}{\partial \mathrm{vec}^T(\mathbf{X})} = (\mathbf{G}^T \otimes \mathbf{I}_m)\frac{\partial \mathrm{vec}[\mathbf{F}(\mathbf{X})]}{\partial \mathrm{vec}^T(\mathbf{X})} + (\mathbf{I}_q \otimes \mathbf{F})\frac{\partial \mathrm{vec}[\mathbf{G}(\mathbf{X})]}{\partial \mathrm{vec}^T(\mathbf{X})}.$$

The commutation matrix $\mathbf{K}_{pm} \in \mathbb{R}^{pm\times pm}$ is the unique matrix that transforms the vec of a matrix into the vec of its transpose: For $\mathbf{F} \in \mathbb{R}^{p\times m}$, $\mathrm{vec}(\mathbf{F}^T) = \mathbf{K}_{pm}\mathrm{vec}(\mathbf{F})$. The next lemma gives properties of commutation matrices used in our derivations.

Lemma D.2 *The following properties hold:*

1. $\mathbf{K}_{pm}^T = \mathbf{K}_{mp}$
2. $\mathbf{K}_{pm}^T\mathbf{K}_{pm} = \mathbf{K}_{pm}\mathbf{K}_{pm}^T = \mathbf{I}_{pm}$;
3. *Suppose* $\mathbf{A} \in \mathbb{R}^{r_1\times r_2}$, $\mathbf{B} \in \mathbb{R}^{r_3\times r_4}$. *Then* $\mathbf{K}_{r_3r_1}(\mathbf{A}\otimes\mathbf{B})\mathbf{K}_{r_2r_4} = \mathbf{B}\otimes\mathbf{A}$.
4. *Suppose* $\mathbf{A} \in \mathbb{S}^{r\times r}$ *and* $\mathbf{C}_r$ *and* $\mathbf{K}_{rr}$ *are defined by* $\mathrm{vech}(\mathbf{A}) = \mathbf{C}_r\mathrm{vec}(\mathbf{A})$ *and* $\mathrm{vec}(\mathbf{A}) = \mathbf{K}_{rr}\mathrm{vec}(\mathbf{A}^T)$. *Then* $\mathbf{C}_r\mathbf{K}_{rr} = \mathbf{C}_r$.

The first term on the right of (S4.4) can now be written as follows:

$$\begin{aligned}
\frac{\partial \mathbf{h}_{21}}{\partial \boldsymbol{\phi}_2^T} &= \frac{\mathbf{C}_r \partial \mathrm{vec}[(\boldsymbol{\Gamma}\boldsymbol{\Omega})\boldsymbol{\Gamma}^T]}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} \\
&= \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \mathbf{I}_r)\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\Omega})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} + \mathbf{C}_r(\mathbf{I}_r \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega})\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} \\
&= \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \mathbf{I}_r)(\boldsymbol{\Omega} \otimes \mathbf{I}_r)\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} + \mathbf{C}_r(\mathbf{I}_r \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega})\mathbf{K}_{ru}\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} \\
&= \mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_r) + \mathbf{C}_r\mathbf{K}_{rr}(\mathbf{I}_r \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega})\mathbf{K}_{ru} \\
&= 2\mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_r). \qquad \text{(S4.5)}
\end{aligned}$$

By a similar derivation, we have

$$\frac{\partial \mathbf{h}_{22}}{\partial \boldsymbol{\phi}_2} = \frac{\partial \mathrm{vech}(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} = 2\mathbf{C}_r(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0 \otimes \mathbf{I}_r)\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}_0)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})}. \qquad \text{(S4.6)}$$

To complete this derivative, we need to define $\boldsymbol{\Gamma}_0$ so that it is uniquely associated with $\boldsymbol{\Gamma}$. One way to do so is the following. First we find it helpful to deviate temporarily from convention and make a careful distinction between $\boldsymbol{\Gamma}$ which can vary and its true fixed population value. Let the columns of the semi-orthogonal matrix $\boldsymbol{\Lambda}$ represent any fixed basis for $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ and let $\boldsymbol{\Lambda}_0$ be any fixed basis for the orthogonal complement of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$. Then in the vicinity of $\boldsymbol{\Lambda}$, $\mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Lambda}_0$ and $\mathbf{Q}_{\boldsymbol{\Gamma}}$ share the same column space. Thus we can take $\mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Lambda}_0$ as the uniquely determined $\boldsymbol{\Gamma}_0$, and the required derivative becomes

$$\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}_0)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} = \frac{\partial \mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Lambda}_0)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} = -\frac{\partial \mathrm{vec}(\mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Lambda}_0)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} = -(\boldsymbol{\Lambda}_0^T \otimes \mathbf{I}_r)\frac{\partial \mathrm{vec}(\mathbf{P}_{\boldsymbol{\Gamma}})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})}$$

The next derivative that we need is

$$\begin{aligned}
\frac{\partial \mathrm{vec}(\mathbf{P}_{\boldsymbol{\Gamma}})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} &= \frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} \\
&= (\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1} \otimes \mathbf{I}_r)\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} + (\mathbf{I}_r \otimes \boldsymbol{\Gamma})\frac{\partial \mathrm{vec}((\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})}
\end{aligned}$$

The first term on the right hand side is 0 when multiplied by $(\boldsymbol{\Lambda}_0^T \otimes \mathbf{I}_r)$. For the second term:

$$\begin{aligned}
\frac{\partial \mathrm{vec}((\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} =& (\boldsymbol{\Gamma} \otimes \mathbf{I}_u)\frac{\partial \mathrm{vec}((\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} + (\mathbf{I}_r \otimes (\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1})\frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}^T)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} \\
=& (\boldsymbol{\Gamma} \otimes \mathbf{I}_u)\frac{\partial \mathrm{vec}((\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1})}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} + (\mathbf{I}_r \otimes (\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1})\mathbf{K}_{ru}
\end{aligned}$$

The first term on the right hand side is again 0 when multiplied by $(\boldsymbol{\Lambda}_0^T \otimes \mathbf{I}_r)$, and therefore

$$\frac{\partial \mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Lambda}_0)}{\partial \mathrm{vec}^T(\boldsymbol{\Gamma})} = -(\boldsymbol{\Lambda}_0^T \otimes \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1})\mathbf{K}_{ru} = -(\boldsymbol{\Lambda}_0^T \otimes \boldsymbol{\Lambda})\mathbf{K}_{ru}$$

where the final term is explicitly evaluated at the true values, recalling that $\boldsymbol{\Lambda}^T\boldsymbol{\Lambda} = \mathbf{I}_u$

Substituting back and returning to the original notation, in which it is understood that all derivatives are evaluated at the true population values, we obtain

$$\begin{aligned}\frac{\partial \mathbf{h}_2}{\partial \boldsymbol{\phi}_2^T} &= 2\mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_r) - 2\mathbf{C}_r(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0 \otimes \mathbf{I}_r)(\boldsymbol{\Gamma}_0^T \otimes \boldsymbol{\Gamma})\mathbf{K}_{ru} \\ &= 2\mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_r - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T).\end{aligned}$$

Finally, we calculate $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}_3^T$ and $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}_4^T$. We have

$$\begin{aligned}\frac{\partial \mathbf{h}_2}{\partial \boldsymbol{\phi}_3^T} &= \frac{\partial \mathrm{vech}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T)}{\partial \mathrm{vech}^T(\boldsymbol{\Omega})} \\ &= \mathbf{C}_r \frac{\partial \mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T)}{\partial \mathrm{vech}^T(\boldsymbol{\Omega})} \\ &= \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_u(\partial \mathrm{vech}(\boldsymbol{\Omega})/\partial \mathrm{vech}^T(\boldsymbol{\Omega})) \\ &= \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_u.\end{aligned}$$

Similarly, $\partial \mathbf{h}_2/\partial \boldsymbol{\phi}_4^T = \mathbf{C}_r(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\mathbf{E}_{(r-u)}$. Now assemble these derivatives together to obtain (24). ■

Proof of (27) Lemma D.2 implies the following corollaries, which will be used repeatedly in the subsequent development.

Corollary D.1 *Let $\mathbf{P}_{\mathbf{E}_r}$ be the projection $\mathbf{E}_r(\mathbf{E}_r^T\mathbf{E}_r)^{-1}\mathbf{E}_r^T$, and $\mathbf{A}$ be an $r \times u$ matrix. Then the following relations hold:*

1. $\mathbf{E}_r\mathbf{C}_r = (\mathbf{E}_r\mathbf{C}_r)^T = \frac{1}{2}(\mathbf{I}_{r^2} + \mathbf{K}_{rr}) = \mathbf{P}_{\mathbf{E}_r}$;
2. $\mathbf{E}_r\mathbf{C}_r(\mathbf{A} \otimes \mathbf{A})\mathbf{E}_u = (\mathbf{A} \otimes \mathbf{A})\mathbf{E}_u$ *and* $\mathbf{E}_r\mathbf{C}_r(\mathbf{A} \otimes \mathbf{A})\mathbf{C}_u^T = (\mathbf{A} \otimes \mathbf{A})\mathbf{C}_u^T$.
3. $\mathbf{P}_{\mathbf{E}_r}(\mathbf{A} \otimes \mathbf{A})\mathbf{P}_{\mathbf{E}_u} = \mathbf{P}_{\mathbf{E}_r}(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A})\mathbf{P}_{\mathbf{E}_u}$;
4. *If* $\mathbf{B} \in \mathbb{R}^{t \times u}$, *then* $\mathbf{K}_{tr}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{A} \otimes \mathbf{B})\mathbf{K}_{su}$.

Corollary D.2 *Let* $\mathbf{C} \in \mathbb{R}^{s \times r}$, $\mathbf{D} \in \mathbb{R}^{t \times r}$, $\mathbf{A} \in \mathbb{R}^{r \times u}$ *and* $\mathbf{B} \in \mathbb{R}^{r \times v}$, *then*

$$(\mathbf{C} \otimes \mathbf{D})\mathbf{P}_{\mathbf{E}_r}(\mathbf{A} \otimes \mathbf{B}) = \frac{1}{2}(\mathbf{CA} \otimes \mathbf{DB}) + \frac{1}{2}(\mathbf{CB} \otimes \mathbf{DA})\mathbf{K}_{vu}.$$

In particular, if either $\mathbf{CB} = \mathbf{0}$ *or* $\mathbf{DA} = \mathbf{0}$, *then*

$$(\mathbf{C} \otimes \mathbf{D})\mathbf{P}_{\mathbf{E}_r}(\mathbf{A} \otimes \mathbf{B}) = \frac{1}{2}(\mathbf{CA} \otimes \mathbf{DB}).$$

Moreover, the following equalities follows from Corollary 2.1.

Corollary D.3

$$\mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma}_0 = \mathbf{\Omega}_0^{-1},\ \ \mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma}_0 = \mathbf{0},\ \ \mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma} = \mathbf{\Omega}^{-1}. \tag{S4.7}$$

We now derive (27). Straightforward matrix multiplication yields

$$\begin{aligned}&\mathbf{H}_{12}^T\mathbf{J}\mathbf{H}_{12} = \boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T \otimes \mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma}_0\\ &+2(\mathbf{\Omega}\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}_0^T - \mathbf{\Gamma}^T \otimes \mathbf{\Omega}_0\mathbf{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\mathbf{\Gamma}\mathbf{\Omega}\otimes\mathbf{\Gamma}_0 - \mathbf{\Gamma}\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0).\end{aligned}$$

By Corollary D.1, part 3, the factor $\mathbf{C}_r^T\mathbf{E}_r^T$ in the second term on the right can be removed. Hence the second term reduces to

$$2(\mathbf{\Omega}\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1} \otimes \mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1} - \mathbf{\Gamma}^T\mathbf{\Sigma}^{-1} \otimes \mathbf{\Omega}_0\mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_r}(\mathbf{\Gamma}\mathbf{\Omega}\otimes\mathbf{\Gamma}_0 - \mathbf{\Gamma}\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0). \tag{S4.8}$$

This can be expanded as 4 terms and, using (S4.7), it can be easily verified that each of these terms is of the form $(\mathbf{C}\otimes\mathbf{D})\mathbf{P}_{\mathbf{E}_r}(\mathbf{A}\otimes\mathbf{B})$ with either $\mathbf{CB} = \mathbf{0}$ or $\mathbf{DA} = \mathbf{0}$. Hence, by Corollary D.2, we can replace $\mathbf{P}_{\mathbf{E}_r}$ by 1/2 in (S4.8), which then reduces to

$$(\mathbf{\Omega}\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1} \otimes \mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1} - \mathbf{\Gamma}^T\mathbf{\Sigma}^{-1} \otimes \mathbf{\Omega}_0\mathbf{\Gamma}_0^T\mathbf{\Sigma}^{-1})(\mathbf{\Gamma}\mathbf{\Omega}\otimes\mathbf{\Gamma}_0 - \mathbf{\Gamma}\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0).$$

Now simplify this using (S4.7) to complete the proof. ∎

E Proofs for Section 6

Block-matrices in (30) From the definitions of $\mathbf{H}$ and $\mathbf{J}$ we have

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} =&(\mathbf{I}_p\otimes\mathbf{\Gamma}^T)(\mathbf{\Sigma}_{\mathbf{X}}\otimes\mathbf{\Sigma}^{-1})(\mathbf{I}_p\otimes\mathbf{\Gamma}) = \mathbf{\Sigma}_{\mathbf{X}}\otimes\mathbf{\Omega}^{-1}\\ \mathbf{J}_{\boldsymbol{\eta}\mathbf{\Gamma}} =&(\mathbf{\Sigma}_{\mathbf{X}}\otimes\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1})(\boldsymbol{\eta}^T\otimes\mathbf{I}_r) = \mathbf{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T\otimes\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}\\ \mathbf{J}_{\mathbf{\Gamma}\mathbf{\Gamma}} =&\boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T\otimes\mathbf{\Sigma}^{-1}\\ &+2(\mathbf{\Omega}\mathbf{\Gamma}^T\otimes\mathbf{I}_r - \mathbf{\Gamma}^T\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\mathbf{\Gamma}\mathbf{\Omega}\otimes\mathbf{I}_r - \mathbf{\Gamma}\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T)\\ \mathbf{J}_{\mathbf{\Gamma}\mathbf{\Omega}} =&(\mathbf{\Omega}\mathbf{\Gamma}^T\otimes\mathbf{I}_r - \mathbf{\Gamma}^T\otimes\mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\mathbf{E}_u\\ =&(\mathbf{\Omega}\mathbf{\Gamma}^T\mathbf{\Sigma}^{-1}\mathbf{\Gamma}\otimes\mathbf{I}_r\mathbf{\Sigma}^{-1}\mathbf{\Gamma})\mathbf{E}_u\\ =&(\mathbf{I}_u\otimes\mathbf{\Sigma}^{-1}\mathbf{\Gamma})\mathbf{E}_u.\\ \mathbf{J}_{\mathbf{\Omega}\mathbf{\Omega}} =&\frac{1}{2}\mathbf{E}_u^T(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\mathbf{C}_r^T\mathbf{E}_r^T(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\mathbf{E}_u = \frac{1}{2}\mathbf{E}_u^T(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1})\mathbf{E}_u\\ \mathbf{J}_{\mathbf{\Omega}_0\mathbf{\Omega}_0} =&\frac{1}{2}\mathbf{E}_{r-u}^T(\mathbf{\Gamma}_0^T\otimes\mathbf{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\mathbf{E}_{r-u}\\ =&\frac{1}{2}\mathbf{E}_{r-u}^T(\mathbf{\Omega}_0^{-1}\otimes\mathbf{\Omega}_0^{-1})\mathbf{E}_{r-u}.\end{aligned}$$

In the derivations of $\mathbf{J}_{\boldsymbol{\Gamma\Omega}}$, $\mathbf{J}_{\boldsymbol{\Omega\Omega}}$, and $\mathbf{J}_{\boldsymbol{\Omega}_0\boldsymbol{\Omega}_0}$ we have used Corollary D.1, part 2, to remove $\mathbf{E}_r\mathbf{C}_r$ and $\mathbf{C}_r^T\mathbf{E}_r^T$ in various places. ■

Derivation of (32) We need the following result, which is a direct consequence of Corollary 2.1 and Proposition 2.2.

Corollary E.1 *Let* $\mathbf{A}$ *be an* $r \times r$ *symmetric, nonsingular matrix and* $\mathbf{G}$ *be an* $r \times u$ *matrix with* $u \leq r$, *assume that* $\mathbf{G}$ *has full column rank. If* $\mathbf{P}_{\mathbf{G}}$ *and* $\mathbf{A}$ *commute, then*

$$\mathbf{G}(\mathbf{G}^T\mathbf{A}\mathbf{G})^{-1}\mathbf{G}^T = \mathbf{P}_{\mathbf{G}}\mathbf{A}^{-1}\mathbf{P}_{\mathbf{G}}.$$

We now derive (32). The first equality in (32) holds because, as we have argued in Section 6, $\text{avar}(\hat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma}}) = \mathbf{J}_{\boldsymbol{\eta\eta}}^{-1}$, which is the desired matrix by the formula for $\mathbf{J}_{\boldsymbol{\eta\eta}}$ given previously in this section.

By (31) and the formulas in Section E, $[\text{avar}(\sqrt{n}\text{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))]^{-1}$ is

$$\begin{aligned}
&\boldsymbol{\eta}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T \otimes \boldsymbol{\Sigma}^{-1} \\
&+2(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T \otimes \mathbf{I}_r - \boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{E}_r\mathbf{C}_r(\boldsymbol{\Gamma\Omega} \otimes \mathbf{I}_r - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) \\
&-2(\mathbf{I}_u \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma})\mathbf{E}_u\left[\mathbf{E}_u^T(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\mathbf{E}_u\right]^{-1}\mathbf{E}_u^T(\mathbf{I}_u \otimes \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1}).
\end{aligned} \tag{S5.9}$$

The second term in (S5.9), without the proportionality constant 2, can be decomposed into the following 4 terms

$$\begin{aligned}
&(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T \otimes \mathbf{I}_r)\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Gamma\Omega} \otimes \mathbf{I}_r) \\
&-(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T \otimes \mathbf{I}_r)\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) \\
&-(\boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Gamma\Omega} \otimes \mathbf{I}_r) \\
&+(\boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) \equiv \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 + \mathbf{A}_4.
\end{aligned} \tag{S5.10}$$

The terms $\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$ can be simplified by the method used in the proof of (27). That is, we can replace the $\mathbf{P}_{\mathbf{E}_u}$ in these terms by 1/2, which results in

$$\mathbf{A}_2 = \mathbf{A}_3 = -\frac{1}{2}\mathbf{I}_u \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T, \quad \mathbf{A}_4 = \frac{1}{2}\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T.$$

Next, let us simplify the first term in (S5.10). By Corollary D.2,

$$(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Gamma\Omega} \otimes \mathbf{I}_r) = \frac{1}{2}(\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1}) + \frac{1}{2}(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma\Omega})\mathbf{K}_{ru}.$$

Let us now simplify the third term in (S5.9). By Corollary D.1, part 3, $\mathbf{P}_{\mathbf{E}_u}$ and $\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}$ commute. Hence, by Corollary E.1, the third term in (S5.9), without the proportionality constant -2, is $(\mathbf{I}_u \otimes$

$\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma})\mathbf{P}_{\mathbf{E}_u}(\boldsymbol{\Omega}\otimes\boldsymbol{\Omega})\mathbf{P}_{\mathbf{E}_u}(\mathbf{I}_u\otimes\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1})$, which, by Corollary D.2, reduces to $\{(\boldsymbol{\Omega}\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T)+(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Omega})\mathbf{K}_{ru}\}/2$. To summarize, we have

$$[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))]^{-1} =\boldsymbol{\eta}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T\otimes\boldsymbol{\Sigma}^{-1}+(\boldsymbol{\Omega}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T)-2(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T)+\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T.$$

This proves the second equality in (32). ■

Proof of Theorem 6.1 Let $\mathbf{H}_{1,ij}$, $i=1,2$, $j=1,\ldots,4$, denote the (i,j)th block of the matrix $\mathbf{H}_1$ defined by (25). Then the asymptotic variance of $\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\beta}})=\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\eta}})$ is

$$\sum_{j=1}^{4}\mathbf{H}_{1,1j}(\mathbf{H}_{1j}^T\mathbf{J}\mathbf{H}_{1j})^{-1}\mathbf{H}_{1,1j}^T.$$

Because $\mathbf{H}_{1,1j}=0$ for $j=3,4$, we have

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\beta}})]=\mathbf{H}_{1,11}(\mathbf{H}_{11}^T\mathbf{J}\mathbf{H}_{11})^{-1}\mathbf{H}_{1,11}^T+\mathbf{H}_{1,12}(\mathbf{H}_{12}^T\mathbf{J}\mathbf{H}_{12})^{-1}\mathbf{H}_{1,12}^T. \tag{S5.11}$$

From the definitions of $\mathbf{H}_1$ and $\mathbf{J}$ we see that

$$\begin{aligned}\mathbf{H}_{11}^T\mathbf{J}\mathbf{H}_{11} =&\boldsymbol{\Sigma}_{\mathbf{X}}\otimes\boldsymbol{\Omega}^{-1}\\ \mathbf{H}_{12}^T\mathbf{J}\mathbf{H}_{12} =&\boldsymbol{\eta}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\eta}^T\otimes\boldsymbol{\Omega}_0^{-1}\\ &+2(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}_0^T-\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{H}_r^T\mathbf{E}_r^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{E}_r\mathbf{H}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega}\otimes\boldsymbol{\Gamma}_0-\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0).\end{aligned}$$

Hence

$$\mathbf{H}_{1,11}(\mathbf{H}_{11}^T\mathbf{J}\mathbf{H}_{11})^{-1}\mathbf{H}_{1,11}^T=(\mathbf{I}_p\otimes\boldsymbol{\Gamma})(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Omega})(\mathbf{I}_p\otimes\boldsymbol{\Gamma}^T)=(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T).$$

Comparing the right hand side with the first equality in (32), we see that

$$\mathbf{H}_{1,11}(\mathbf{H}_{11}^T\mathbf{J}\mathbf{H}_{11})^{-1}\mathbf{H}_{1,11}^T=(\mathbf{I}_p\otimes\boldsymbol{\Gamma})\mathrm{avar}[\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma}})](\mathbf{I}_p\otimes\boldsymbol{\Gamma}^T). \tag{S5.12}$$

In the meantime, comparing (27) and the second equality in (32) we see that

$$\mathbf{H}_{12}\mathbf{J}\mathbf{H}_{12}^T=(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0^T)[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))]^{-1}(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0).$$

Consequently,

$$\begin{aligned}&\mathbf{H}_{1,12}(\mathbf{H}_{12}\mathbf{J}\mathbf{H}_{12}^T)^{-1}\mathbf{H}_{1,12}^T\\ =&(\boldsymbol{\eta}^T\otimes\mathbf{I}_r)(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0)\left\{(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0^T)[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))]^{-1}(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0)\right\}^{-1}(\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0^T)(\boldsymbol{\eta}\otimes\mathbf{I}_r).\end{aligned}$$

Now let the $\mathbf{G}$ and $\mathbf{A}$ in Lemma E.1 be $\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0$ and $[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))]^{-1}$, respectively. Then $\mathbf{P}_{\mathbf{G}}=\mathbf{I}_u\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T$, and it is easy to verify that $\mathbf{P}_{\mathbf{G}}$ and $\mathbf{A}$ commute. Hence

$$\begin{aligned}\mathbf{H}_{1,12}(\mathbf{H}_{12}\mathbf{J}\mathbf{H}_{12}^T)^{-1}\mathbf{H}_{1,12}^T =&(\boldsymbol{\eta}^T\otimes\mathbf{I}_r)(\mathbf{I}_u\otimes\mathbf{P}_{\boldsymbol{\Gamma}_0})[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))](\mathbf{I}_u\otimes\mathbf{P}_{\boldsymbol{\Gamma}_0})(\boldsymbol{\eta}\otimes\mathbf{I}_r)\\ =&(\boldsymbol{\eta}^T\otimes\mathbf{P}_{\boldsymbol{\Gamma}_0})[\mathrm{avar}(\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta}}))](\boldsymbol{\eta}\otimes\mathbf{P}_{\boldsymbol{\Gamma}_0}).\end{aligned} \tag{S5.13}$$

Now substitute (S5.12) and (S5.13) into (S5.11) to complete the proof. ■