ENVELOPE MODELS FOR PARSIMONIOUS AND EFFICIENT MULTIVARIATE LINEAR REGRESSION

R. Dennis Cook\textsuperscript{1}, Bing Li\textsuperscript{2} and Francesca Chiaromonte\textsuperscript{2}

\textsuperscript{1}University of Minnesota and \textsuperscript{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 \( M \in \mathbb{R}^{r \times r} \). \( \mathcal{R} \) is an invariant subspace of \( M \) if and only if, for any \( A \in \mathbb{R}^{r \times s} \), \( s \geq u \), such that span\((A) = \mathcal{R} \), there exists a \( B \in \mathbb{R}^{s \times s} \) such that \( MA = AB \).

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

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

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

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

Assume that \( \mathcal{R} \subseteq \text{span}(M) \). Then it follows from Corollary 2.1 that \( A^TMA \) is of full rank. Thus, \( B \) must have full rank in the representation \( MA = AB \), which implies \( MR = \mathcal{R} \). □
Lemma S1.3  Suppose that \( \mathcal{R} \) reduces \( M \in \mathbb{S}^{r \times r} \). Then \( M \) has a spectral decomposition with eigenvectors in \( \mathcal{R} \) or in \( \mathcal{R}^\perp \).

Proof. Let \( A_0 \in \mathbb{R}^{r \times u} \) be a semi-orthogonal matrix whose columns span \( \mathcal{R} \) and let \( A_1 \) be its completion, such that \((A_0, A_1) = A\) is an orthogonal matrix. Because \( \mathcal{MR} \subseteq \mathcal{R} \) and \( \mathcal{MR}^\perp \subseteq \mathcal{R}^\perp \), it follows from Lemma S1.1 there exist matrices \( B_0 \in \mathbb{R}^{u \times u} \) and \( B_1 \in \mathbb{R}^{(r-u) \times (r-u)} \) such that \( MA_0 = A_0B_0 \) and \( MA_1 = A_1B_1 \). Hence

\[
M = A_0B_0\begin{pmatrix} A_0 & A_1 \end{pmatrix} = \begin{pmatrix} B_0 & 0 \\ 0 & B_1 \end{pmatrix} \iff M = A \begin{pmatrix} B_0 & 0 \\ 0 & B_1 \end{pmatrix} A^T.
\]

Because \( M \) is symmetric, so must be \( B_0 \) and \( B_1 \). Hence \( B_0 \) and \( B_1 \) have spectral decompositions \( C_0A_0C_0^T \) and \( C_1A_1C_1^T \), for some diagonal matrices \( A_0 \) and \( A_1 \) and orthogonal matrices \( C_0 \) and \( C_1 \). Let \( C = \text{diag}(C_0, C_1) \) and \( A = \text{diag}(A_0, A_1) \). Then,

\[
M = ACAC^T A^T = DAD^T,
\]

where \( D = AC \). The first \( u \) columns of \( D \), which form the matrix \( A_0C_0 \), span \( \mathcal{R} \). Moreover, \( D \) is an orthogonal matrix, and thus (S1.1) is a spectral decomposition of \( M \) with eigenvectors in \( \mathcal{R} \) or \( \mathcal{R}^\perp \). ■

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

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

\[
M = (P_\mathcal{R} + Q_\mathcal{R})M(P_\mathcal{R} + Q_\mathcal{R}) = P_\mathcal{R}MP_\mathcal{R} + Q_\mathcal{R}MQ_\mathcal{R}.
\]

■

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

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

For the third conclusion, since \( M \) is full rank \( \mathcal{R} \subseteq \text{span}(M) \) and \( \mathcal{R}^\perp \subseteq \text{span}(M) \). Consequently, both \( A^TMA \) and \( A_0^TMA_0 \) are full rank. Thus the right hand side of (5) is defined. Meanwhile, note that \( P_\mathcal{R} = AA^T \) and \( Q_\mathcal{R} = A_0A_0^T \). Hence, by (4), \( M = AA^TMAA^T + A_0A_0^TMA_0A_0^T \). 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 \( v \in R \), then

\[
v = I_p v = \left( \sum_{i=1}^{q} P_i \right) v = \sum_{i=1}^{q} P_i v \in \oplus_{i=1}^{q} P_i R.
\]

Hence \( R \subseteq \oplus_{i=1}^{q} P_i R \). Conversely, if \( v \in \oplus_{i=1}^{q} P_i R \), then \( v \) can be written as a linear combination of \( P_1 v_1, \ldots, P_q v_q \) where \( v_1, \ldots, v_q \) belong to \( R \). By Lemma S1.3, \( P_i w \in R \) for any \( w \in R \). Hence any linear combination of \( P_1 v_1, \ldots, P_q v_q \), with \( v_1, \ldots, v_q \) belonging to \( R \), belongs to \( R \). That is, \( \oplus_{i=1}^{q} P_i R \subseteq R \).

2 implies 3: If \( v \in R \) then, from the previous step, \( P_i v \in R \), \( i = 1, \ldots, q \). Hence

\[
\left( \sum_{i=1}^{q} P_i P_R P_i \right) v = \sum_{i=1}^{q} P_i v = v = P_R v.
\]

Now let \( v \in R^\perp \). Then, \( v \perp P_i R \) for each \( i \). Because \( P_i \) is self-adjoint we see that \( P_i v \perp R \) for each \( i \). Consequently

\[
\left( \sum_{i=1}^{q} P_i P_R P_i \right) v = 0 = P_R v.
\]

It follows that \( (\sum P_i P_R P_i) v = P_R v \) for all \( v \in R^\perp \). Hence the two matrices are the same.

3 implies 1: Again, if \( v \in R \) then \( P_i v \in R \), \( i = 1, \ldots, q \). Hence, indicating with \( m_i \), \( i = 1, \ldots, q \) the distinct eigenvalues of \( M \) we have

\[
P_R M v = \sum_{i=1}^{q} m_i P_i P_R P_i P_R P_i v = \sum_{i=1}^{q} m_i P_i P_R P_i v = M v.
\]

It follows that \( M R \subseteq R \).

Proof of Proposition 2.3  To prove that \( \oplus_{i=1}^{q} P_i S \) is the smallest reducing subspace of \( M \) that contains \( S \), it suffices to prove the following statements:

1. \( \oplus_{i=1}^{q} P_i S \) reduces \( M \).

2. \( S \subseteq \oplus_{i=1}^{q} P_i S \).

3. If \( T \) reduces \( M \) and \( S \subseteq T \), then \( \oplus_{i=1}^{q} P_i S \subseteq T \).
Statement 1 follows from Proposition 2.2, as applied to \( R \equiv \oplus_{i=1}^q P_i S \). Statement 2 holds because \( S = \{ P_1 v + \cdots + P_q v : v \in S \} \subseteq \oplus_{i=1}^q P_i S \). Turning to statement 3, if \( T \) reduces \( M \), it can be written as \( T = \oplus_{i=1}^q P_i T \) by Proposition 2.2. If, in addition, \( S \subseteq T \) then we have \( P_i S \subseteq P_i T \) for \( i = 1, \ldots, q \). Statement 3 follows since \( \oplus_{i=1}^q P_i S \subseteq \oplus_{i=1}^q P_i T = T \).

Proof of Proposition 2.4 Because \( K \) and \( M \) commute, they can be diagonalized simultaneously by an orthogonal matrix, say \( U \). Recall that \( P_i \) is the projection on the \( i \)-th eigenspace of \( M \), and let \( d_i = \text{rank}(P_i) \). Partition \( U = (U_1, \ldots, U_q) \), where \( U_i \) contains \( d_i \) columns, \( i = 1, \ldots, q \). Without loss of generality, we can assume that \( U_i U_i^T = P_i \) for \( i = 1, \ldots, q \). Then \( K \) can be written as \( U_1 \Lambda_1 U_1^T + \cdots + U_q \Lambda_q U_q^T \), where the \( \Lambda_i \)'s are diagonal matrices of dimension \( d_i \times d_i \). It follows that

\[
KP_i = (U_1 \Lambda_1 U_1^T + \cdots + U_q \Lambda_q U_q^T) U_i U_i^T = U_i \Lambda_i U_i^T(U_1 \Lambda_1 U_1^T + \cdots + U_q \Lambda_q U_q^T) = P_i K.
\]

That is, \( K \) and \( P_i \) commute. Now, by Proposition 2.2, \( \mathcal{E}_M(S) = \oplus_{i=1}^q P_i S \). Hence

\[
K \mathcal{E}_M(S) = \{ KP_1 h_1 + \cdots + KP_q h_q : h_1, \ldots, h_q \in S \} = \{ P_1 K h_1 + \cdots + P_q K h_q : h_1, \ldots, h_q \in S \} = \oplus_{i=1}^q P_i K S.
\]

By Proposition 2.2 again, the right hand side is \( \mathcal{E}_M(KS) \).

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

S2 Proofs for Section 3

Proof of Proposition 3.1 We need to show that \( \Sigma^{-1} B = \Sigma^{-1} B \), and

\[
\mathcal{E}_\Sigma(B) = \mathcal{E}_\Sigma(B) = \mathcal{E}_{\Sigma}(\Sigma^{-1} B) = \mathcal{E}_{\Sigma}(\Sigma^{-1} B) = \mathcal{E}_\Sigma(\Sigma^{-1} B) = \mathcal{E}_\Sigma(\Sigma^{-1} B).
\]

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

\[
\Sigma_Y^{-1} = \Sigma^{-1} - \Sigma^{-1} \Gamma (\Gamma^{-1} + \Gamma^T \Sigma^{-1} \Gamma) \Gamma^T \Sigma^{-1},
\]

\[
\Sigma^{-1} = \Sigma_Y^{-1} - \Sigma_Y^{-1} \Gamma (-\Gamma^{-1} + \Gamma^T \Sigma_Y^{-1} \Gamma) \Gamma^T \Sigma_Y^{-1}.
\]
The first equality implies \( \text{span}(\Sigma_Y^{-1}\Gamma) \subseteq \text{span}(\Sigma^{-1}\Gamma) \); the second implies \( \text{span}(\Sigma_Y^{-1}\Gamma) \subseteq \text{span}(\Sigma_Y^{-1}\Gamma) \). Hence \( \Sigma_Y^{-1}B = \Sigma^{-1}B \), recalling that \( \text{span}(\Gamma) = \mathcal{E}_\Sigma(B) \) by construction. From this we also deduce \( \mathcal{E}_{\Sigma_Y}(\Sigma_Y^{-1}B) = \mathcal{E}_{\Sigma_Y}(\Sigma^{-1}B) \) and \( \mathcal{E}_{\Sigma_Y}(\Sigma_Y^{-1}B) = \mathcal{E}_{\Sigma}(\Sigma^{-1}B) \).

We next show that \( \mathcal{E}_{\Sigma}(B) = \mathcal{E}_{\Sigma_Y}(B) \) by demonstrating that \( R \subseteq \mathbb{R}^p \) is a reducing subspace of \( \Sigma \) that contains \( B \) if and only if it is a reducing subspace of \( \Sigma_Y \) that contains \( B \). Suppose \( R \subseteq \mathbb{R}^p \) is a reducing subspace of \( \Sigma_Y \) and by construction it contains \( B \). The reverse implication follows similarly by reasoning in terms of \( \Sigma \).

The remaining equalities follow immediately from (8).

\[ \blacksquare \]

**S3 Proofs for Section 4**

**Proof of Lemma 4.1** From the properties of a projection we see that, for any \( B \in A \), we have \( P_{W(A)}B^TP_V = B^T \). It follows that

\[
\text{tr}(A^*AB^T) = \text{tr}[P_VUP_{W(A)}AB^T] = \text{tr}[UAP_{W(A)}B^TP_V] = \text{tr}(UB^T).
\]

Thus \( \text{tr}[(U - A^*)AB] = 0 \). Now decompose the objective function (11) as

\[
\]

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

\[
\text{tr}[(U - A)\Lambda(U - A)^T] = \text{tr}[(U - A^*)\Lambda(A^* - A)^T] + \text{tr}[(A^* - A)\Lambda(A^* - A)^T] 
\]

\[
\geq \text{tr}[(U - A^*)\Lambda(U - A^*)^T].
\]

The lower bound is achieved when \( A = A^* \), in which case

\[
\text{tr}[(U - A^*)\Lambda(U - A^*)^T] = \text{tr}[(U - A^*)\Lambda U^T] = \text{tr}([U\Lambda U^T] - (A^*\Lambda U^T)).
\]

However, by definition of \( A^* \) and the property of projection,

\[
\text{tr}(A^*\Lambda U^T) = \text{tr}(P_VUP_{W(A)}\Lambda U^T) = \text{tr}[P_VUP_{W(A)}\Lambda P_{W(A)}U^TP_V],
\]

as desired. \[ \blacksquare \]

**Proof of Lemma 4.1** Straightforward and omitted.
Proof of Lemma 4.3  Because $P$ is the projection onto $\text{span}(A)$, we have

$$L(A) = [\det_0(A)]^{-\frac{1}{2}}e^{-\frac{1}{2}[\text{tr}(UP\Gamma^T)]}. \quad (S3.2)$$

If we write $U^T = (U_1, \ldots, U_n)$, then the above is proportional to the likelihood of $U_1, \ldots, U_n$ if they are iid $N(0, A)$. The maximum likelihood estimator (among all $A$) is the sample variance $A^* = PU^TUP/n$. Note that $A^*$ happens to be in $\mathcal{A}$. Therefore $L(A)$ is maximized by $A^*$ among $A$. In the meantime,

$$\text{tr}[UP(A^*)\Gamma^T] = \text{tr}[(A^*)\Gamma^TUP] = n\text{tr}[(A^*)\Gamma^T] = nk,$$

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

S4 Proofs for Section 5

Proof of Theorem 5.1 The asymptotic distribution (23), where $H$ is as defined in (22), follows from Shapiro (1986, Proposition 4.1). To prove the equality $V_0 \leq V$ we note that

$$V - V_0 = J^{-1} - H(H^TJH)^{\frac{1}{2}}H^T = J^{-\frac{1}{2}}[I_{pr+r(r+1)/2} - J^{\frac{1}{2}}H(H^TJH)^{\frac{1}{2}}H^T]J^{\frac{1}{2}}J^{-\frac{1}{2}}. \quad (S4.3)$$

Since the matrix $I_{pr+r(r+1)/2} - J^{\frac{1}{2}}H(H^TJH)^{\frac{1}{2}}H^TJ^{\frac{1}{2}}$ is the projection on to orthogonal complement of $\text{span}(J^{\frac{1}{2}}H)$ relative to the standard inner product, it is positive semidefinite, which implies that $V - V_0$ is positive semidefinite. From (S4.3) we can also see that

$$V^{-\frac{1}{2}}(V - V_0)V^{-\frac{1}{2}} = I_{pr+r(r+1)/2} - J^{\frac{1}{2}}H(H^TJH)^{\frac{1}{2}}H^TJ^{\frac{1}{2}} = QJ^{\frac{1}{2}}H^T,$$

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

Step 1: Compute $\partial h_1/\partial \phi_1^T$

First,

$$\frac{\partial h_1}{\partial \phi_1^T} = \frac{\partial \text{vec}(\Gamma\eta)}{\partial \text{vec}(\eta)} = \frac{\partial[(I_p \otimes \Gamma)\text{vec}(\eta)]}{\partial \text{vec}(\eta)} = I_p \otimes \Gamma \in \mathbb{R}^{pr \times pu}.$$

In a similar way,

$$\frac{\partial h_1}{\partial \phi_2^T} = \frac{\partial \text{vec}(\Gamma\eta)}{\partial \text{vec}(\Gamma)} = \eta^T \otimes I_r \in \mathbb{R}^{pr \times ur}.$$
Clearly, $\partial h_1/\partial \phi_T^3 = 0$, $\partial h_1/\partial \phi_T^4 = 0$, 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 $\partial h_2/\partial \phi_T^T$

Since $h_2$ does not depend on $\phi_1$ we have $\partial h_2/\partial \phi_T^1 = 0$. Note that this matrix is of dimension $r(r+1)/2 \times pu$.

To compute $\partial h_2/\partial \phi_T^2$, let $h_{21}(\phi) = \text{vech}(\Gamma \Omega \Gamma^T)$ and $h_{22}(\phi) = \text{vech}(\Gamma_0 \Omega_0 \Gamma_0^T)$, so that we can write

$$\frac{\partial h_2}{\partial \phi_T^2} = \frac{\partial h_{21}}{\partial \phi_T^2} + \frac{\partial h_{22}}{\partial \phi_T^2}. \quad (S4.4)$$

where $\phi_2 = \text{vec}(\Gamma)$. The following two lemmas, which are presented without proof, will facilitate computation of the derivatives in (S4.4).

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

$$\frac{\partial \text{vec}[F(X)G(X)]}{\partial \text{vec}^T(X)} = (G^T \otimes I_m) \frac{\partial \text{vec}[F(X)]}{\partial \text{vec}^T(X)} + (I_q \otimes F) \frac{\partial \text{vec}[G(X)]}{\partial \text{vec}^T(X)}.$$ 

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

**Lemma D.2** The following properties hold:

1. $K_{pm}^T = K_{mp}$

2. $K_{pm}K_{pm} = K_{pm}K_{pm} = I_{pm}$;

3. Suppose $A \in \mathbb{R}^{r_1 \times r_2}$, $B \in \mathbb{R}^{r_3 \times r_4}$. Then $K_{r_3 r_1}(A \otimes B)K_{r_2 r_4} = B \otimes A$.

4. Suppose $A \in \mathbb{R}^{r \times r}$ and $C_r$ and $K_{rr}$ are defined by vech($A$) = $C_r$vec($A$) and vec($A$) = $K_{rr}$vec($A^T$). Then $C_rK_{rr} = C_r$. 


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

\[
\frac{\partial h_{21}}{\partial \phi_2} = \frac{C_r \partial \text{vec}([\Gamma \Omega]) \Gamma^T}{\partial \text{vec}^T(\Gamma)} = C_r(\Gamma \otimes I_r) \frac{\partial \text{vec}(\Gamma \Omega)}{\partial \text{vec}^T(\Gamma)} + C_r(I_r \otimes \Gamma \Omega) \frac{\partial \text{vec}(\Gamma^T)}{\partial \text{vec}^T(\Gamma)}
\]

\[
= C_r(\Gamma \otimes I_r) \partial \text{vec}(\Gamma \Omega) + C_r(I_r \otimes \Gamma \Omega) \partial \text{vec}(\Gamma^T)
\]

\[
= C_r(\Gamma \Omega \otimes I_r) + C_r(K_{rr}(I_r \otimes \Gamma \Omega))K_{ru}
\]

\[
= 2C_r(\Gamma \Omega \otimes I_r).
\]  

(S4.5)

By a similar derivation, we have

\[
\frac{\partial h_{22}}{\partial \phi_2} = \frac{\partial \text{vec}(\Gamma_0 \Omega_0 \Gamma_0^T)}{\partial \text{vec}^T(\Gamma)} = 2C_r(\Gamma_0 \Omega_0 \otimes I_r) \frac{\partial \text{vec}(\Gamma_0)}{\partial \text{vec}^T(\Gamma)}.
\]  

(S4.6)

To complete this derivative, we need to define \( \Gamma_0 \) so that it is uniquely associated with \( \Gamma \). One way to do so is the following. First we find it helpful to deviate temporarily from convention and make a careful distinction between \( \Gamma \) which can vary and its true fixed population value. Let the columns of the semi-orthogonal matrix \( \Lambda \) represent any fixed basis for \( \mathcal{E}_B(B) \) and let \( \Lambda_0 \) be any fixed basis for the orthogonal complement of \( \mathcal{E}_B(B) \). Then in the vicinity of \( \Lambda \), \( Q_{fr} \Lambda_0 \) and \( Q_{fr} \) share the same column space. Thus we can take \( Q_{fr} \Lambda_0 \) as the uniquely determined \( \Gamma_0 \), and the required derivative becomes

\[
\frac{\partial \text{vec}(\Gamma_0)}{\partial \text{vec}^T(\Gamma)} = \frac{\partial \text{vec}(Q_{fr} \Lambda_0)}{\partial \text{vec}^T(\Gamma)} = \frac{\partial \text{vec}(P_{fr} \Lambda_0)}{\partial \text{vec}^T(\Gamma)} = -\left( \Lambda_0^T \otimes I_r \right) \frac{\partial \text{vec}(P_{fr})}{\partial \text{vec}^T(\Gamma)}
\]

The next derivative that we need is

\[
\frac{\partial \text{vec}(P_{fr})}{\partial \text{vec}^T(\Gamma)} = \frac{\partial \text{vec}(\Gamma(\Gamma^T \Gamma)^{-1} \Gamma^T)}{\partial \text{vec}^T(\Gamma)}
\]

\[
= (\Gamma(\Gamma^T \Gamma)^{-1} \otimes I_r) \frac{\partial \text{vec}(\Gamma)}{\partial \text{vec}^T(\Gamma)} + (I_r \otimes \Gamma) \frac{\partial \text{vec}((\Gamma^T \Gamma)^{-1} \Gamma^T)}{\partial \text{vec}^T(\Gamma)}
\]

The first term on the right hand side is 0 when multiplied by \( (\Lambda_0^T \otimes I_r) \). For the second term:

\[
\frac{\partial \text{vec}((\Gamma^T \Gamma)^{-1} \Gamma^T)}{\partial \text{vec}^T(\Gamma)} = (\Gamma \otimes I_u) \frac{\partial \text{vec}((\Gamma^T \Gamma)^{-1})}{\partial \text{vec}^T(\Gamma)} + (I_r \otimes (\Gamma^T \Gamma)^{-1}) \frac{\partial \text{vec}(\Gamma^T)}{\partial \text{vec}^T(\Gamma)}
\]

\[
= (\Gamma \otimes I_u) \frac{\partial \text{vec}((\Gamma^T \Gamma)^{-1})}{\partial \text{vec}^T(\Gamma)} + (I_r \otimes (\Gamma^T \Gamma)^{-1})K_{ru}
\]

The first term on the right hand side is again 0 when multiplied by \( (\Lambda_0^T \otimes I_r) \), and therefore

\[
\frac{\partial \text{vec}(Q_{fr} \Lambda_0)}{\partial \text{vec}^T(\Gamma)} = -\left( \Lambda_0^T \otimes \Gamma(\Gamma^T \Gamma)^{-1} \right)K_{ru} = -\left( \Lambda_0^T \otimes \Lambda \right)K_{ru}
\]
where the final term is explicitly evaluated at the true values, recalling that $A^T A = I_u$

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

$$
\frac{\partial h_2}{\partial \phi^T_2} = 2C_r(\Gamma \Omega \otimes I_r) - 2C_r(\Gamma_0 \Omega_0 \otimes I_r)(\Gamma_0^T \otimes \Gamma)K_{ru}
$$

Finally, we calculate $\frac{\partial h_2}{\partial \phi^T_3}$ and $\frac{\partial h_2}{\partial \phi^T_4}$. We have

$$
\frac{\partial h_2}{\partial \phi^T_3} = \frac{\partial \text{vech}(\Gamma \Omega \Gamma^T)}{\partial \text{vech}^T(\Omega)} = \frac{\partial \text{vech}(\Gamma_0 \Omega_0 \Gamma_0^T)}{\partial \text{vech}^T(\Omega)}
$$

Similarly, $\frac{\partial h_2}{\partial \phi^T_4} = C_r(\Gamma_0 \otimes \Gamma_0)E_{(r-u)}$. 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 $P_{E_r}$ be the projection $E_r(E_r^T E_r)^{-1}E_r^T$, and $A$ be an $r \times u$ matrix. Then the following relations hold:

1. $E_r C_r = (E_r C_r)^T = \frac{1}{2}(I_r + K_{rr}) = P_{E_r}$;

2. $E_r A E_u = (A \otimes A)E_u$ and $E_r (A \otimes A)C_u^T = (A \otimes A)C_u^T$.

3. $P_{E_r} (A \otimes A)P_{E_u} = P_{E_r} (A \otimes A) = (A \otimes A)P_{E_u}$;

4. If $B \in \mathbb{R}^{t \times u}$, then $K_{tr}(A \otimes B) = (A \otimes B)K_{su}$.

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

$$(C \otimes D)P_{E_r} (A \otimes B) = \frac{1}{2}(CA \otimes DB) + \frac{1}{2}(CB \otimes DA)K_{vu}.$$ 

In particular, if either $CB = 0$ or $DA = 0$, then

$$(C \otimes D)P_{E_r} (A \otimes B) = \frac{1}{2}(CA \otimes DB).$$
Moreover, the following equalities follows from Corollary 2.1.

**Corollary D.3**

\[
\Gamma_0^T \Sigma^{-1} \Gamma_0 = \Omega_0^{-1}, \quad \Gamma_0^T \Sigma^{-1} \Gamma_0 = 0, \quad \Gamma^T \Sigma^{-1} \Gamma = \Omega^{-1}.
\]  
\[
(S4.7)
\]

We now derive (27). Straightforward matrix multiplication yields

\[
H_{12}^T J H_{12} = \eta \Sigma \eta^T \Sigma^{-1} \Gamma_0
\]

\[
+ 2(\Omega \Gamma^T \Sigma^{-1} \Gamma_0 - \Gamma^T \Sigma^{-1} \Gamma_0 \Gamma^T \Sigma^{-1} \Gamma_0) C_r \Gamma_0^T (\Sigma^{-1} \Omega \Sigma^{-1}) E_r C_r (\Gamma \Sigma^{-1} \Gamma_0 - \Gamma \Gamma_0 \Omega_0).
\]

By Corollary D.1, part 3, the factor \( C_r \Gamma_0^T E_r \) in the second term on the right can be removed. Hence the second term reduces to

\[
2(\Omega \Gamma^T \Sigma^{-1} \Gamma_0 - \Gamma^T \Gamma_0 \Gamma^T \Sigma^{-1} \Gamma_0) P_E (\Gamma \Omega_0 - \Gamma \Gamma_0 \Omega_0).
\]

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 \((C \otimes D)P_E, (A \otimes B)\) with either \(CB = 0\) or \(DA = 0\). Hence, by Corollary D.2, we can replace \(P_E\) by 1/2 in (S4.8), which then reduces to

\[
(\Omega \Gamma^T \Sigma^{-1} \Gamma_0 - \Gamma^T \Gamma_0 \Gamma^T \Sigma^{-1} \Gamma_0)(\Gamma \Omega_0 - \Gamma \Gamma_0 \Omega_0).
\]

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

---

### E Proofs for Section 6

**Block-matrices in (30)** From the definitions of \(H\) and \(J\) we have

\[
J_{\eta \eta} = (I_p \otimes \Gamma^T)(\Sigma \otimes \Sigma^{-1})(I_p \otimes \Gamma) = \Sigma \otimes \Omega^{-1}
\]

\[
J_{\eta \Gamma} = (\Sigma \otimes \Gamma^T \Sigma^{-1})(\eta^T \otimes I_r) = \Sigma \eta^T \otimes \Gamma^T \Sigma^{-1}
\]

\[
J_{\Gamma \Gamma} = \eta \Sigma \eta^T \otimes \Sigma^{-1}
\]

\[
+ 2(\Omega \Gamma^T \otimes I_r - \Gamma^T \otimes \Gamma_0 \Omega_0 \Gamma^T \Sigma^{-1} \Gamma_0) C_r \Gamma_0^T (\Sigma^{-1} \Omega \Sigma^{-1}) E_r C_r (\Gamma \Omega_0 \otimes I_r - \Gamma \Gamma_0 \Omega_0 \Gamma^T \Sigma^{-1} \Gamma_0)
\]

\[
J_{\Gamma \Omega} = (\Omega \Gamma^T \otimes I_r - \Gamma^T \otimes \Gamma_0 \Omega_0 \Gamma^T \Sigma^{-1} \Gamma_0) C_r \Gamma_0^T (\Sigma^{-1} \Omega \Sigma^{-1}) E_r C_r (\Gamma \Omega_0 \otimes I_r - \Gamma \Gamma_0 \Omega_0 \Gamma^T \Sigma^{-1} \Gamma_0)
\]

\[
= (\Omega \Gamma^T \Sigma^{-1} \Gamma_0 \otimes I_r \Sigma^{-1} \Gamma_0) E_u
\]

\[
= (I_u \otimes \Sigma^{-1} \Gamma_0 \otimes \Sigma^{-1} \Gamma_0) E_u.
\]

\[
J_{\Omega \Omega} = \frac{1}{2} E_u (\Gamma^T \otimes \Gamma^T \otimes \Gamma^T \Sigma^{-1} \Omega \Sigma^{-1}) E_r C_r (\Gamma \otimes \Gamma) E_u = \frac{1}{2} E_u (\Omega^{-1} \otimes \Omega^{-1}) E_u
\]

\[
J_{\Omega \Omega_0} = \frac{1}{2} E_u (\Gamma_0^T \otimes \Gamma_0^T \otimes \Gamma_0^T \Sigma^{-1} \Omega \Sigma^{-1}) E_r C_r (\Gamma \otimes \Gamma) E_{r-u} = \frac{1}{2} E_u (\Omega_0^{-1} \otimes \Omega_0^{-1}) E_{r-u}.
\]
In the derivations of $J_{\Omega\Omega}$, $J_{\Omega\Omega}$, and $J_{\Omega\Omega\Omega}$, we have used Corollary D.1, part 2, to remove $E_rC_r$ and $C_r^TE_r^T$ 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 $A$ be an $r \times r$ symmetric, nonsingular matrix and $G$ be an $r \times u$ matrix with $u \leq r$, assume that $G$ has full column rank. If $P_G$ and $A$ commute, then

$$G(G^T AG)^{-1}G^T = P_G A^{-1}P_G.$$  

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

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

$$\eta \Sigma \times \eta^T \otimes \Sigma^{-1}$$

$$+ 2(\Omega \Gamma^T \otimes I_r - \Gamma^T \otimes \Gamma_0 \Omega_0 \Gamma_0^T)C_r^T E_r^T (\Sigma^{-1} \otimes \Sigma^{-1})E_r C_u (\Omega \Gamma^T \otimes I_r - \Gamma \otimes \Gamma_0 \Omega_0 \Gamma_0^T)$$

$$- 2(I_u \otimes \Sigma^{-1} \Gamma)E_u [E_u^T (\Omega^{-1} \otimes \Omega^{-1})E_u]^{-1} E_u^T (I_u \otimes \Gamma^T \Sigma^{-1}).$$

(S5.9)

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

$$= (\Omega \Gamma^T \otimes I_r)P_{E_u}(\Sigma^{-1} \otimes \Sigma^{-1})P_{E_u}(\Omega \Gamma^T \otimes I_r)$$

$$- (\Omega \Gamma^T \otimes I_r)P_{E_u}(\Sigma^{-1} \otimes \Sigma^{-1})P_{E_u}(\Gamma \otimes \Gamma_0 \Omega_0 \Gamma_0^T)$$

$$- (\Gamma^T \otimes \Gamma_0 \Omega_0 \Gamma_0^T)P_{E_u}(\Sigma^{-1} \otimes \Sigma^{-1})P_{E_u}(\Gamma \otimes \Gamma_0 \Omega_0 \Gamma_0^T)$$

$$+ (\Gamma^T \otimes \Gamma_0 \Omega_0 \Gamma_0^T)P_{E_u}(\Sigma^{-1} \otimes \Sigma^{-1})P_{E_u}(\Gamma \otimes \Gamma_0 \Omega_0 \Gamma_0^T) \equiv A_1 + A_2 + A_3 + A_4.$$  

(S5.10)

The terms $A_2, A_3, A_4$ can be simplified by the method used in the proof of (27). That is, we can replace the $P_{E_u}$ in these terms by $1/2$, which results in

$$A_2 = A_3 = -\frac{1}{2} I_u \otimes \Gamma_0 \Gamma_0^T, \quad A_4 = \frac{1}{2} \Omega^{-1} \otimes \Gamma_0 \Omega_0 \Gamma_0^T.$$  

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

$$(\Omega \Gamma^T \Sigma^{-1} \otimes \Sigma^{-1})P_{E_u}(\Omega \Gamma^T \otimes I_r) = \frac{1}{2} (\Omega \otimes \Sigma^{-1}) + \frac{1}{2} (\Omega \Gamma^T \Sigma^{-1} \otimes \Sigma^{-1} \Gamma \Omega) K_{ru}.$$  

Let us now simplify the third term in (S5.9). By Corollary D.1, part 3, $P_{E_u}$ and $\Omega^{-1} \otimes \Omega^{-1}$ commute. Hence, by Corollary E.1, the third term in (S5.9), without the proportionality constant $-2$, is $(I_u \otimes$
\[ \Sigma^{-1} \Gamma \] \[ P_{E_\nu} (\Omega \otimes \Xi) \] \[ P_{E_\nu} (I_u \otimes \Gamma^T \Sigma^{-1}) \], which, by Corollary D.2, reduces to \( \{ (\Omega \otimes \Gamma \Omega^{-1} \Gamma^T) + (\Omega \Gamma^T \Sigma^{-1} \otimes \Sigma^{-1} \Gamma \Omega) K_{\tau \nu} \}/2 \). To summarize, we have

\[ [\text{avar}(\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta))]^{-1} = \eta \Sigma^\chi \eta^T \otimes \Sigma^{-1} + (\Omega \otimes \Gamma_0 \Omega_0^{-1} \Gamma_0^T) - 2(I_u \otimes \Gamma_0 \Gamma_0^T) + \Omega^{-1} \otimes \Gamma_0 \Omega_0 \Gamma_0^T. \]

This proves the second equality in (32).

**Proof of Theorem 6.1** Let \( H_{1,ij}, i = 1, 2, j = 1, \ldots, 4 \), denote the \((i, j)\)th block of the matrix \( H_1 \) defined by (25). Then the asymptotic variance of \( \sqrt{n} \text{vec}(\tilde{\beta}) = \sqrt{n} \text{vec}(\tilde{\Gamma}_\eta) \) is

\[ \sum_{j=1}^{4} H_{1,1j} (H_{1j}^T J H_{1j})^{-1} H_{1,1j}^T. \]

Because \( H_{1,1j} = 0 \) for \( j = 3, 4 \), we have

\[ \text{avar}[\sqrt{n} \text{vec}(\tilde{\beta})] = H_{1,11} (H_{11}^T J H_{11})^{-1} H_{1,11}^T + H_{1,12} (H_{12}^T J H_{12})^{-1} H_{1,12}. \]  

(S5.11)

From the definitions of \( H_1 \) and \( J \) we see that

\[ H_{11}^T J H_{11} = \Sigma^\chi \otimes \Omega^{-1} \]

\[ H_{12}^T J H_{12} = \eta \Sigma^\chi \eta^T \otimes \Omega_0^{-1} + 2(\Omega \Gamma^T \otimes \Gamma_0^T - \Gamma^T \otimes \Omega_0 \Gamma_0^T) H_1^T E_r (\Sigma^{-1} \otimes \Sigma^{-1}) E_r H_1 (\Gamma \Omega \otimes \Gamma_0 - \Gamma \otimes \Gamma_0 \Omega_0). \]

Hence

\[ H_{1,11} (H_{11}^T J H_{11})^{-1} H_{1,11}^T = (I_p \otimes \Gamma)(\Sigma^\chi^{-1} \otimes \Omega)(I_p \otimes \Gamma^T) = (\Sigma^\chi^{-1} \otimes \Gamma \Omega \Gamma^T). \]

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

\[ H_{1,11} (H_{11}^T J H_{11})^{-1} H_{1,11}^T = (I_p \otimes \Gamma) \text{avar}[\sqrt{n} \text{vec}(\tilde{\eta}_\tau)] (I_p \otimes \Gamma^T). \]  

(S5.12)

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

\[ H_{12} J H_{12}^T = (I_u \otimes \Gamma_0^T) \text{avar}[\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta)]^{-1} (I_u \otimes \Gamma_0). \]

Consequently,

\[ H_{1,12} (H_{12} J H_{12}^T)^{-1} H_{1,12}^T \]

\[ = (\eta^T \otimes I_r) (I_u \otimes \Gamma_0) \left\{ (I_u \otimes \Gamma_0^T) \text{avar}[\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta)]^{-1} (I_u \otimes \Gamma_0) \right\}^{-1} (I_u \otimes \Gamma_0^T) (\eta \otimes I_r). \]

Now let the \( G \) and \( A \) in Lemma E.1 be \( I_u \otimes \Gamma_0 \) and \( \text{avar}[\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta)]^{-1} \), respectively. Then \( P_{G} = I_u \otimes \Gamma_0 \Gamma_0^T \), and it is easy to verify that \( P_{G} \) and \( A \) commute. Hence

\[ H_{1,12} (H_{12} J H_{12}^T)^{-1} H_{1,12}^T = (\eta^T \otimes I_r) (I_u \otimes P_{G_0}) \text{avar}[\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta)] (I_u \otimes P_{G_0}) (\eta \otimes I_r) \]

\[ = (\eta^T \otimes P_{G_0}) \text{avar}[\sqrt{n} \text{vec}(\tilde{\Gamma}_\eta)] (\eta \otimes P_{G_0}). \]  

(S5.13)

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

\[ \blacksquare \]