New Parsimonious Multivariate Spatial Model: Spatial Envelope

S1 Brief Review of Linear Coregionalization Model

Linear Coregionalization Models (LCM) is popular in multivariate spatial data analysis. This model assumes that the observed variables are linear combinations of sets of independent underlying variables and they covary jointly over a region. Various methods have been proposed for fitting LCM in literatures including least square approach (Goulard and Voltz, 1992) and expectation maximization (EM) algorithm (Zhang, 2007), among others. Let $\mathbf{Y}(s_i) = (y_1(s_i), \ldots, y_r(s_i))^T$ be an r-variate stochastic spatial response vector along with p regressors $X(s_i) = (x_1(s_i), \ldots, x_p(s_i))^T$ observed at locations $s = \{s_1, s_2, \ldots, s_n;\ s_i \in \mathbb{R}^2; i = 1, 2, \ldots, n\}$. The most basic version of the LCM can be written as

$$\mathbf{Y}(s_i) = \boldsymbol{\mu} + \sum_{k=0}^{K} \mathbf{W}_k(s_i), \tag{S1.1}$$

where $\boldsymbol{\mu}$ denotes the mean. The mean term, $\boldsymbol{\mu}$, shows the trend in data and it is common to use a linear model of covariates such as $\boldsymbol{\alpha} + \boldsymbol{\beta} X(s_i)$ to model this term. In model (S1.1), $\mathbf{W}_0(s_i)$ is a stationary but uncorrelated r-variate process with mean 0 and multivariate covariance function $\Sigma_0 = \mathbf{V}_0$. Furthermore, for $k = 1, \ldots, K$, $\mathbf{W}_k(s_i)$ are i.i.d spatial processes with mean 0 and multivariate covariance function $\Sigma_k = \mathbf{V}_k \rho_k(\mathbf{h})$ where $\mathbf{h} = ||s_i - s_j||$ denotes the Euclidean distance between location s_i and s_j. The log likelihood function of model (S1.1) can be written as

$$\begin{aligned}
\log L(\Delta, \mathbf{Y}, \mathbf{W}) &= -\frac{1}{2}\log\det(\boldsymbol{\Sigma}_0) \\
&- \frac{1}{2}\left(\mathbf{Y} - \boldsymbol{\mu}\otimes\mathbf{1} - \sum_{k=1}^{K}\mathbf{W}_k\right)^T \boldsymbol{\Sigma}_0^{-1}\left(\mathbf{Y} - \boldsymbol{\mu}\otimes\mathbf{1} - \sum_{k=1}^{K}\mathbf{W}_k\right) \\
&- \frac{1}{2}\sum_{k=1}^{K}\left(\log(\det(\Sigma_k)) + \mathbf{W}_k^T\Sigma_k^{-1}\mathbf{W}_k\right),
\end{aligned}$$

where $\mathbf{Y} = (\mathbf{Y}^T(s_1), \ldots, \mathbf{Y}^T(s_n))^T$ denotes the reponse, $\Delta = (\boldsymbol{\mu}, \mathbf{V}_0, \mathbf{V}_k, \rho_k)$; $k = 1, \ldots, K$, denotes the parameters in the model, $\otimes$ denotes the Kronecker product, and $\mathbf{1}$ is a vector which all of its entries are 1. In addition, for the spatial processes, let $\mathbf{W}_{ki} = (\mathbf{W}_{ki}(s_1), \ldots, \mathbf{W}_{ki}(s_n))^T$, $\mathbf{W}_k = (\mathbf{W}_{k1}^T, \ldots, \mathbf{W}_{kp}^T)^T$, and $\mathbf{W} = (\mathbf{W}_0^T, \ldots, \mathbf{W}_0^r)^T$. Since in the LCM, the $\mathbf{W}$ are unobserved, the expectation-maximization (EM) algorithm can be applied to estimate the parameters. For further information on the EM algorithm for maximum-likelihood estimation of the LCM, see Zhang (2007) and the references therein.

S2 Derivation of the factorization of the likelihood function in section 4.1

The likelihood function of the model (3.6) will be as follows:

$$L^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta}) = [\det((\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{1}{2}}$$
$$\times\, exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)^T \left((\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}(\boldsymbol{\theta})\right)^{-1} (\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)\right\}$$
$$= [\det(\mathbf{V}_0 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}) + \mathbf{V}_1 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{1}{2}}$$
$$\times\, exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)^T \left((\mathbf{V}_0 + \mathbf{V}_1)^{-1} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right) (\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)\right\}$$
$$= [\det(\mathbf{V}_0 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}) + \mathbf{V}_1 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{1}{2}}$$
$$\times\, exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)^T \left((\mathbf{V}_0^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})) + (\mathbf{V}_1^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\right) (\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)\right\}, \tag{S2.1}$$

where † denotes Moore-Penrose inverse and $\mathbf{V}_0 = \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0$ and $\mathbf{V}_1 = \boldsymbol{\Gamma}_1\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_1$. Since $span(\boldsymbol{\beta}) \subseteq span(\mathbf{V}_1)$ and $\boldsymbol{\beta} = \boldsymbol{\Gamma}_1\boldsymbol{\eta}$, therefore we have $\boldsymbol{\beta}^T = \boldsymbol{\eta}^T\boldsymbol{\Gamma}_1^T$ which means

$$\boldsymbol{\beta}^* = vec(\boldsymbol{\beta}^T) = vec(\boldsymbol{\eta}^T\boldsymbol{\Gamma}_1^T) = (\boldsymbol{\Gamma}_1 \otimes \boldsymbol{\eta}^T)vec(\mathbf{I}_u).$$

Last equality holds by the results of theorem 11.6a in Seber (2008). Thus we have

$$
\begin{aligned}
(\mathbf{V}_0^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\mathbb{X}\boldsymbol{\beta}^* &= (\mathbf{V}_0^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))(\mathbf{I}_r \otimes \mathbf{X})\boldsymbol{\beta}^* \\
&= (\mathbf{V}_0^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))(\mathbf{I}_r \otimes \mathbf{X})(\boldsymbol{\Gamma}_1 \otimes \boldsymbol{\eta}^T)vec(\mathbf{I}_u) \\
&= (\mathbf{V}_0^{\dagger}\boldsymbol{\Gamma}_1 \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\eta}^T)vec(\mathbf{I}_u) \\
&= (\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T\boldsymbol{\Gamma}_1 \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\eta}^T)vec(\mathbf{I}_u) \\
&= \mathbf{0},
\end{aligned}
$$

the last equality holds because $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ are orthagonal. Therefore, Since $(\mathbf{V}_0^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\mathbb{X}\boldsymbol{\beta}^* = \mathbf{0}$ and $\mathbf{V} = \mathbf{V}_0 + \mathbf{V}_1$, the likelihood in (S2.1) can be factored as:

$$
\begin{aligned}
& L^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta}) = [\det((\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}(\boldsymbol{\theta})) \\
& \times exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)^T \left(\mathbf{V}_1^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)\right\} \\
& \times exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n)^T \left(\mathbf{V}_0^{\dagger} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n)\right\} \\
& = L_1^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_1, \boldsymbol{\theta}) \times L_2^u(\boldsymbol{\alpha}, \mathbf{V}_0, \boldsymbol{\theta}),
\end{aligned}
\tag{S2.2}
$$

where

$$
\begin{aligned}
L_1^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_1, \boldsymbol{\theta}) &= [det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} \\
&\times exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)^T \left(\mathbf{V}_1^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right) (\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}\boldsymbol{\beta}^*)\right\}, \\
L_2^u(\boldsymbol{\alpha}, \mathbf{V}_0, \boldsymbol{\theta}) &= [det_0(\mathbf{V}_0)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} \\
&\times exp\left\{-\frac{1}{2}(\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n)^T \left(\mathbf{V}_0^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right) (\mathbb{Y} - \boldsymbol{\alpha} \otimes \mathbf{1}_n)\right\},
\end{aligned} \tag{S2.3}
$$

where $det_0(\mathbf{A})$ denotes the product of non-zero eigenvalues of $\mathbf{A}$ where $\mathbf{A}$ is a non-zero symmetric matrix. This is due to

$$
\begin{aligned}
\det((\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}(\boldsymbol{\theta})) &= \det[\mathbf{V}_0 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}) + \mathbf{V}_1 \otimes \boldsymbol{\rho}(\boldsymbol{\theta})] \\
&= det_0[\mathbf{V}_0 \otimes \boldsymbol{\rho}(\boldsymbol{\theta})] + det_0[\mathbf{V}_1 \otimes \boldsymbol{\rho}(\boldsymbol{\theta})] \\
&= [det_0(\mathbf{V}_0)]^n[det_0(\boldsymbol{\rho}(\boldsymbol{\theta}))]^r + [det_0(\mathbf{V}_1)]^n[det_0(\boldsymbol{\rho}(\boldsymbol{\theta}))]^r \\
&= [det_0(\mathbf{V}_0)]^n[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^r + [det_0(\mathbf{V}_1)]^n[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^r
\end{aligned}
$$

the last equality holds because is $\boldsymbol{\rho}(\boldsymbol{\theta})$ a full rank positive definite matrix therefore $det_0 = \det$.

S3 Coordinate free version of the algorithm of the spatial envelope

The objective is to maximize the likelihood in (3.7) over $\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_0, \mathbf{V}_1$, and $\boldsymbol{\theta}$ subject to the constraints:

$$\begin{aligned} &span(\boldsymbol{\beta}) \subseteq span(\mathbf{V}_1), \quad (a) \\ &\mathbf{V}_0\mathbf{V}_1 = 0, \qquad\qquad\qquad (b). \end{aligned} \tag{S3.1}$$

Based on this factorization given in equation (S2.2), we can decompose the likelihood maximization into the following steps:

1. Fix $\boldsymbol{\beta}, \mathbf{V}_0$, $\mathbf{V}_1$, and $\boldsymbol{\theta}$, and maximize $L^{(u)}$ in (3.6) over $\boldsymbol{\alpha}$ which will be:

$$\hat{\boldsymbol{\alpha}} = \bar{\mathbf{Y}} - \bar{\mathbf{X}}\boldsymbol{\beta}^T.$$

Let $\mathbf{H} = \mathbf{Y} - \bar{\mathbf{Y}} \otimes \mathbf{1}_n$, $\mathbf{U} = vec(\mathbf{H})$, $\mathbf{G} = \mathbf{X} - \bar{\mathbf{X}} \otimes \mathbf{1}_n$, and $\mathbf{F} = \mathbf{I}_r \otimes \mathbf{G}$. Therefore, the profile likelihood can be written as the following:

$$\begin{aligned} L_1^u(\boldsymbol{\beta}^*, \mathbf{V}_1, \boldsymbol{\theta}) &= [det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} \\ &\times exp\left\{-\frac{1}{2}(\mathbf{U} - \mathbf{F}\boldsymbol{\beta}^*)^T \left(\mathbf{V}_1^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)(\mathbf{U} - \mathbf{F}\boldsymbol{\beta}^*)\right\}, \end{aligned} \tag{S3.2}$$

and

$$L_2^u(\mathbf{V}_0, \boldsymbol{\theta}) = [det_0(\mathbf{V}_0)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} exp\left\{-\frac{1}{2}\mathbf{U}^T \left(\mathbf{V}_0^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)\mathbf{U}\right\}. \tag{S3.3}$$

2. Fix $\mathbf{V}_1$, and $\boldsymbol{\theta}$ and maximize the function L_1^u over $\boldsymbol{\beta}^*$, subject to (S3.1a), to obtain $L_{21}^u(\mathbf{V}_1, \boldsymbol{\theta})$. Since $vec(\mathbf{AB}) = (\mathbf{I}_r \otimes \mathbf{A})vec(\mathbf{B}^T)$ and

$$tr(\mathbf{D}^T(\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T)) = (vec(\mathbf{D}))^T(\mathbf{A} \otimes \mathbf{C}^T)(vec(\mathbf{B}))^T,$$

we have

$$\begin{aligned}
(\mathbf{U} - \mathbf{F}\boldsymbol{\beta}^*)^T \left(\mathbf{V}_1^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)(\mathbf{U} - \mathbf{F}\boldsymbol{\beta}^*) &= tr\left((\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})(\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)\mathbf{V}_1^\dagger\right) \\
&= tr\left((\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})(\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)\mathbf{V}_1^\dagger\right) \\
&= tr\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})(\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)\mathbf{V}_1^\dagger(\mathbf{H} - \mathbf{G}\boldsymbol{\beta}^T)^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\right) \\
&= tr\left(\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}_1^\dagger\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T\right) \\
&= tr\left(\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{I}_r\right)\mathbf{V}_1^\dagger\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{I}_r\right)^T\right)
\end{aligned} \tag{S3.4}$$

where $tr(\cdot)$ denotes the trace of the matrix. The last equality in equation (S3.4) is from Lemma 4.1 in Cook, Li, and Chiaromonte (2010). Thus, the optimal $\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{I}_r$ is

$$\mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{P}^T_{(\mathbf{I}_r(\mathbf{V}_1^\dagger))} = \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\left(\boldsymbol{\rho}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H}\right)\mathbf{P}_{\mathbf{V}_1},$$

where $\mathbf{P}_{(\cdot)}$ is the projection onto the subspace indicated by its argument. This implies following

$$\boldsymbol{\beta}^T = \left(\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right)^{-1}\mathbf{G}\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_1} \Rightarrow \boldsymbol{\beta} = \mathbf{P}_{\mathbf{V}_1}\hat{\boldsymbol{\beta}},$$

where $\boldsymbol{\beta}$ is the MLE estimate of $\boldsymbol{\beta}$ from the full model (3.6). Substituting this into (S3.3) and using the relation $\mathbf{P}_{\mathbf{V}_1}\mathbf{V}_1^\dagger = \mathbf{V}_1^\dagger$, the maximum of $L_2^{(u)}$ for fixed $\mathbf{V}_1$ over $\boldsymbol{\beta}$ is

$$
\begin{aligned}
&L_{11}^u(\mathbf{V}_1,\boldsymbol{\theta}) = [det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} \\
&\times exp\left\{-\frac{1}{2}tr\left(\left(\boldsymbol{\rho}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H} - \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_1}\right)\mathbf{V}_1^{\dagger}\left(\boldsymbol{\rho}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H} - \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_1}\right)^T\right)\right\} \\
&= [det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} \\
&\times exp\left\{-\frac{1}{2}tr\left(\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{V}_1^{\dagger}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)^T\right)\right\} \\
&= [det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-\frac{r}{2}} exp\left\{-\frac{1}{2}tr\left(\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{V}_1^{\dagger}\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)^T\right)\right\}
\end{aligned}
\tag{S3.5}
$$

where $\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)} = \mathbf{I}_n - \mathbf{P}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}$.

3. Maximize $L^u(\mathbf{V}_0,\mathbf{V}_1,\boldsymbol{\theta})$ over all $\mathbf{V}_0$, $\mathbf{V}_1$, and $\boldsymbol{\theta}$. Since $L^u(\mathbf{V}_0,\mathbf{V}_1,\boldsymbol{\theta}) = L_1^u(\mathbf{V}_1,\boldsymbol{\theta}) \times L_2^u(\mathbf{V}_0,\boldsymbol{\theta})$, we have

$$
\begin{aligned}
&L^u(\mathbf{V}_0,\mathbf{V}_1,\boldsymbol{\theta}) = [det_0(\mathbf{V}_0)]^{-\frac{n}{2}}[det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-r} \\
&\times exp\left\{-\frac{1}{2}tr\left(\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{V}_1^{\dagger}\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)^T\right)\right\} \\
&\times exp\left\{-\frac{1}{2}\mathbf{U}^T\left(\mathbf{V}_0^{\dagger}\otimes\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right)\mathbf{U}\right\} \\
&= [det_0(\mathbf{V}_0)]^{-\frac{n}{2}}[det_0(\mathbf{V}_1)]^{-\frac{n}{2}}[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-r} \\
&\times exp\left\{-\frac{1}{2}tr\left(\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{V}_1^{\dagger}\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)^T\right)\right\} \\
&\times exp\left\{-\frac{1}{2}tr\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\mathbf{V}_0^{\dagger}\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\right)\right\}.
\end{aligned}
\tag{S3.6}
$$

This maximization can be as follows:

(a) Fix $\mathbf{V}_0$ and $\mathbf{V}_1$ and maximize $L^u(\mathbf{V}_0,\mathbf{V}_1,\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ by solving the following maximization problem:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\{r \det(\boldsymbol{\rho}(\boldsymbol{\theta})) \\ &+ \frac{1}{2} tr\left(\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{V}_1^{\dagger}\left(\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H}\right)^T + \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\mathbf{V}_0^{\dagger}\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\right)\}.\end{aligned}$$

(b) Fix the $\boldsymbol{\theta}$ and maximize $L^u(\mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta})$ over $\mathbf{V}_0$ and $\mathbf{V}_1$. This means maximize $L_{11}^u(\mathbf{V}_1, \boldsymbol{\theta})$ over $\mathbf{V}_1$ and $L_{12}^u(\mathbf{V}_0, \boldsymbol{\theta})$ over $\mathbf{V}_0$. Maximization $L_{11}^u(\mathbf{P}_{\mathbf{V}_1})$ over $\mathbf{V}_1$ is

$$L_{11}^u(\mathbf{P}_{\mathbf{V}_1}) \propto \left[det_0\left(\mathbf{P}_{\mathbf{V}_1}\left(\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{P}_{\mathbf{V}_1}\right)\right]^{-\frac{n}{2}} \tag{S3.7}$$

and maximization $L_{12}^u(\mathbf{P}_{\mathbf{V}_0})$ over $\mathbf{V}_0$ is

$$L_{12}^u(\mathbf{P}_{\mathbf{V}_0}) \propto \left[det_0\left(\mathbf{P}_{\mathbf{V}_0}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_0}\right)\right]^{-\frac{n}{2}}. \tag{S3.8}$$

Therefore, maximization $L^u(\mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta})$ over $\mathbf{V}_0$ and $\mathbf{V}_1$ is equivalent to maximization of $L_{11}^u(\mathbf{P}_{\mathbf{V}_1}) \times L_{12}^u(\mathbf{P}_{\mathbf{V}_0})$ which is proportion to

$$\begin{aligned}\mathbf{D} &= \left[det_0\left(\mathbf{P}_{\mathbf{V}_1}\left(\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{P}_{\mathbf{V}_1}\right)\right]^{-\frac{n}{2}} \\ &\times \left[det_0\left(\mathbf{P}_{\mathbf{V}_0}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_0}\right)\right]^{-\frac{n}{2}} \\ &= \left[det_0\left(\mathbf{P}_{\mathbf{V}_1}\left(\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{P}_{\mathbf{V}_1} + \mathbf{P}_{\mathbf{V}_0}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{P}_{\mathbf{V}_0}\right)\right]^{-\frac{n}{2}} \\ &= \left[det_0\left(\mathbf{P}_{\mathbf{V}_1}\left(\mathbf{H}^T\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}\right)\mathbf{P}_{\mathbf{V}_1} + \mathbf{Q}_{\mathbf{V}_0}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{Q}_{\mathbf{V}_0}\right)\right]^{-\frac{n}{2}}\end{aligned} \tag{S3.9}$$

where $\mathbf{Q}_{\mathbf{V}_0} = \mathbf{I}_r - \mathbf{P}_{\mathbf{V}_1}$. Since $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} = \mathbf{H}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}$ and

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{res} &= \mathbf{H}^T \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{Q}_{\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\right)}\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} \\ &= \mathbf{H}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H} \\ &- \mathbf{H}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\left(\mathbf{G}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right)^{-1}\mathbf{G}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}.\end{aligned} \tag{S3.10}$$

Therefore we have $\mathbf{D} = det(\mathbf{P}_{\mathbf{V}_1}\hat{\boldsymbol{\Sigma}}_{\mathbf{res}}\mathbf{P}_{\mathbf{V}_1} + \mathbf{Q}_{\mathbf{V}_1}\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}\mathbf{Q}_{\mathbf{V}_1})$ and $\hat{\mathbf{V}}_1 = \text{argmin}_{\mathbf{V}_1}(\mathbf{D})$ and $\mathbf{P}_{\hat{\mathbf{V}}_0} = \mathbf{I}_r - \mathbf{P}_{\hat{\mathbf{V}}_1}$

Repeat (a) and (b) until the difference between estimations of the parameters from two consecutive iterations is smaller than a pre-specified tolerance level.

S4 Proof of Lemma 1

In this section, we derive the Fisher information matrix for the parameters given by equation (4.2). Before starting the derivation, the following properties hold:

1. Suppose $\mathbf{A}$ and $\mathbf{X}$ are both $r \times r$, and $\mathbf{X}$ is symmetric, then

$$\frac{\partial vech(\mathbf{X}^{-1})}{(\partial vech(\mathbf{X}))^T} = -\mathbf{C}_r\left(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}\right)\mathbf{E}_r,$$

where $\mathbf{E}_r \in R^{r^2 \times r(r+1)/2}$ is an expansion matrix such that for a matrix $\mathbf{A}$, $vec(\mathbf{A}) = \mathbf{E}_r vech(\mathbf{A})$, and $\mathbf{C}_r \in R^{r(r+1)/2 \times r^2}$ is expansion matrix which is defined such that for a given matrix such as $\mathbf{A}$, $vech(\mathbf{A}) = \mathbf{C}_r vec(\mathbf{A})$ and $\mathbf{E}_r \in R^{r^2 \times r(r+1)/2}$ is expansion matrix which is defined such that $vec(\mathbf{A}) = \mathbf{E}_r vech(\mathbf{A})$.

2. If $\mathbf{X}$ is nonsingular and unconstrained, then we have

$$\frac{\partial tr(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T.$$

3. If $\mathbf{Y} = \mathbf{AXB}$, then

$$tr(\mathbf{Y}) = vec(\mathbf{A}^T\mathbf{B}^T)vec(X) = vec(\mathbf{A}^T\mathbf{B}^T)\mathbf{E}_n vech(X),$$

and

$$\frac{\partial tr(\mathbf{Y})}{\partial vec(\mathbf{X})} = vec(\mathbf{A}^T\mathbf{B}^T).$$

4. Suppose $\mathbf{B}_1$ is an $m \times n$ and $\mathbf{B}_2$ is an $n \times q$, matrix, then

$$vec(\mathbf{B}_1\mathbf{B}_2) = (\mathbf{B}_2 \otimes \mathbf{I}_m)vec(\mathbf{B}_1).$$

5. Suppose $\mathbf{X}$ is an $m \times n$ and $\mathbf{A}$ is an $n \times n$, matrix, then

$$\frac{\partial vec(\mathbf{XAX})}{\partial (vec(\mathbf{X}))^T} = (\mathbf{X}^T\mathbf{A}^T \otimes \mathbf{I}_n)\mathbf{I}_{nm} + (\mathbf{I}_n \otimes \mathbf{X}^T\mathbf{A}).$$

6. Assume $\mathbf{X}$ to be $m \times n$. Then we have,

$$\frac{\partial(\mathbf{X}^T\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{AX} + \mathbf{A}^T\mathbf{X}.$$

7. Let $\mathbf{P}_{\mathbf{E}_r}$ denotes the projection of $\mathbf{E}_r(\mathbf{E}_r^T\mathbf{E}_r)^{-1}\mathbf{E}_r^T$ then, $\mathbf{P}_{\mathbf{E}_r} = \mathbf{E}_r\mathbf{C}_r$ and $\mathbf{E}_r^T\mathbf{E}_r\mathbf{C}_r = \mathbf{E}_r^T$,

Proof of the first six properties can be found in Seber (2008). The proof of the last property can be found in Cook, Li, and Chiaromonte (2010)

The logarithm of the likelihood function (3.7) is

$$\ell(\boldsymbol{\Theta}) = -\frac{1}{2}\log[\det(\mathbf{V}\otimes\boldsymbol{\rho}(\boldsymbol{\theta}))] - \frac{1}{2}(\mathbb{Y}-\boldsymbol{\alpha}\otimes\mathbf{1}_n-\mathbb{X}\boldsymbol{\beta}^*)^T(\mathbf{V}\otimes\boldsymbol{\rho}(\boldsymbol{\theta}))^{-1}(\mathbb{Y}-\boldsymbol{\alpha}\otimes\mathbf{1}_n-\mathbb{X}\boldsymbol{\beta}^*) \tag{S4.1}$$

where $\boldsymbol{\Theta} = \{\mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}^*, \boldsymbol{\theta}\}$. First and second derivatives of the log likelihood function in (S4.1) with respect to $\boldsymbol{\beta}^*$ are

$$\begin{aligned}
\text{First derivative: } \frac{\partial \ell(\boldsymbol{\Theta})}{\partial \boldsymbol{\beta}^*} &= \mathbb{X}^T(\mathbf{V}^{-1}\otimes\boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))(\mathbb{Y}-\boldsymbol{\alpha}\otimes\mathbf{1}_n-\mathbb{X}\boldsymbol{\beta}^*),\\
\text{Second derivative: } \frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial \boldsymbol{\beta}^*\partial \boldsymbol{\beta}^{*T}} &= -\mathbb{X}^T(\mathbf{V}^{-1}\otimes\boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\mathbb{X}\\
&= -(\mathbf{I}_r\otimes\mathbf{X}^T)(\mathbf{V}^{-1}\otimes\boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))(\mathbf{I}_r\otimes\mathbf{X})\\
&= -\mathbf{V}^{-1}\otimes\left(\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\mathbf{X}\right)
\end{aligned}$$

From (3.7), we can rewrite the log likelihood function as

$$\begin{aligned}
\ell(\boldsymbol{\Theta}) = &-\frac{n}{2}\log[\det(\mathbf{V})] - \frac{r}{2}\log[\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]\\
&-\frac{1}{2}tr\left(\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}-\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}-\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T\right).
\end{aligned} \tag{S4.2}$$

The $tr(\cdot)$ is due to

$$\begin{aligned}
&(\mathbf{U}-\mathbf{F}\boldsymbol{\beta}^*)^T(\mathbf{V}^{-1}\otimes\boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))(\mathbf{U}-\mathbf{F}\boldsymbol{\beta}^*) = tr\left((\mathbf{H}-\mathbf{G}\boldsymbol{\beta}^T)^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})(\mathbf{H}-\mathbf{G}\boldsymbol{\beta}^T)^T\mathbf{V}^{-1}\right)\\
&= tr\left(\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}-\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H}-\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\right).
\end{aligned}$$

Therefore, the first derivative of the log likelihood function in (S4.2) with respect to $\mathbf{V}$ is $\frac{\partial \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})} = \frac{\partial \ell(\boldsymbol{\Theta})}{\partial vec(\mathbf{V})} \frac{\partial vec(\mathbf{V})}{\partial vech(\mathbf{V})}$, where

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})} &= -\frac{n}{2} vec\left(\mathbf{V}^{-1}\right)^T \mathbf{E}_r \\
&+ \frac{1}{2} vec\left\{\mathbf{V}^{-1}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T \left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\right\}\mathbf{E}_r \\
&= -\frac{n}{2} vech\left(\mathbf{V}^{-1}\right)^T \mathbf{E}_r^T \mathbf{E}_r \\
&+ \frac{1}{2} vech\left\{\mathbf{V}^{-1}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T \left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\right\}\mathbf{E}_r^T\mathbf{E}_r
\end{aligned} \tag{S4.3}$$

and second derivative of the log likelihood function in (S4.2) with respect to $\mathbf{V}$ is

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial vech(\mathbf{V})^T} &= \frac{n}{2}\mathbf{E}_r^T\left(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1}\right)\mathbf{E}_r \\
&- \frac{1}{2}\mathbf{A}\mathbf{V}^{-1}\mathbf{E}_r^T(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})\mathbf{C}_r^T\mathbf{E}_r^T\mathbf{E}_r - \frac{1}{2}\mathbf{A}^T\mathbf{V}^{-1}\mathbf{E}_r(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})\mathbf{C}_r^T\mathbf{E}_r^T\mathbf{E}_r
\end{aligned} \tag{S4.4}$$

where $\mathbf{A} = \left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)$. Thus,

$$E\left(\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial vech(\mathbf{V})^T}\right) = -\frac{n}{2}\mathbf{E}_r^T(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})\mathbf{E}_r$$

Finally, we have to calculate $\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial\partial\boldsymbol{\beta}^*\partial vech(\mathbf{V})^T}$ and $\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial\boldsymbol{\beta}^{*T}}$. Since these two are equal, we only calculate the second one.

$$\begin{aligned}
&\frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial\boldsymbol{\beta}^{*T}} = \frac{\partial^2 \ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial(vec(\boldsymbol{\beta}^T))^T} \\
&= \frac{1}{2}\frac{vec\left\{\mathbf{V}^{-1}\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)^T\left(\boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\rho}^{-\frac{1}{2}}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\right\}\mathbf{E}_r}{\partial(vec(\boldsymbol{\beta}^T))^T} \\
&= \frac{1}{2}\frac{vec\left[\mathbf{V}^{-1}\left(\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H} - \boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H} - \mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T + \boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\right)\mathbf{V}^{-1}\right]\mathbf{E}_r}{\partial(vec(\boldsymbol{\beta}^T))^T}.
\end{aligned} \tag{S4.5}$$

The derivative of $vec\left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{V}^{-1}\right)\mathbf{E}_r$ with respect to $vec(\boldsymbol{\beta}^T))^T$ is zero. Furthermore, using matrix algebra, we have

$$\begin{aligned}
vec\left(\mathbf{V}^{-1}\boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{V}^{-1}\right) &= \left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}\mathbf{G}\otimes\mathbf{V}^{-1}\right)vec(\boldsymbol{\beta}) \\
&= \left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}\mathbf{G}\otimes\mathbf{V}^{-1}\right)\mathbf{K}_{rp}vec(\boldsymbol{\beta}^T) \\
vec\left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{V}^{-1}\right) &= \left(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right)vec(\boldsymbol{\beta}^T).
\end{aligned}$$

where $\mathbf{K}_{rp}\in\mathbb{R}^{rp\times rp}$ is the unique matrix that transform the *vec* of a matrix into the *vec* of its transpose i.e. for a given matrix such as $\mathbf{A}\in\mathbb{R}^{m\times n}$ we have $vec(\mathbf{A}^T) = \mathbf{K}_{mn}vec(\mathbf{A})$. More properties of $\mathbf{K}_{mn}$ can be found in Cook, Li, and Chiaromonte (2010) lemma D.2. Therefore, we have

$$\begin{aligned}
\frac{vec\left(\mathbf{V}^{-1}\boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{H}\mathbf{V}^{-1}\right)}{\partial(vec(\boldsymbol{\beta}^T))^T} &= \left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}\mathbf{G}\otimes\mathbf{V}^{-1}\right)\mathbf{K}_{rp} \\
\frac{vec\left(\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{V}^{-1}\right)}{\partial(vec(\boldsymbol{\beta}^T))^T} &= \left(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1}\mathbf{H}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right) \\
\frac{vec\left(\mathbf{V}^{-1}\boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\boldsymbol{\beta}^T\mathbf{V}^{-1}\right)}{\partial(vec(\boldsymbol{\beta}^T))^T} &= \left(\mathbf{V}^{-1}\boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\otimes\mathbf{V}^{-1}\right)\mathbf{K}_{rp} + \left(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1}\boldsymbol{\beta}\mathbf{G}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right).
\end{aligned} \tag{S4.6}$$

Substituting (S4.6) in equation (S4.5), we have

$$\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial\boldsymbol{\beta}^{*T}} &= \frac{1}{2}\left\{\mathbf{V}^{-1}\left(\mathbf{H}-\mathbf{G}\boldsymbol{\beta}^T\right)^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\otimes\mathbf{V}^{-1}\right\}\mathbf{K}_{rp}\mathbf{E}_r \\
&+ \frac{1}{2}\left\{\mathbf{V}^{-1}\otimes\mathbf{V}^{-1}\left(\mathbf{H}-\mathbf{G}\boldsymbol{\beta}^T\right)^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{G}\right\}\mathbf{E}_r
\end{aligned} \tag{S4.7}$$

Taking the expected value of these derivatives together and the fact that

$$E\left[\frac{\partial^2\ell(\boldsymbol{\Theta})}{\partial vech(\mathbf{V})\partial\boldsymbol{\beta}^*}\right] = \mathbf{0},$$

lead to obtain (4.4).

S5 Proof of Theorem 1

In this section, we derive the an explicit expression for $\mathbf{\Psi}$ as given by (4.3). In order to find these expression, we need to find expressions for the eight partial derivatives $\frac{\partial \Psi_i}{\partial \phi_j^T}$ for $i = 1, 2$ and $j = 1, 2, 3, 4$.

Theorem 1: Suppose $\bar{\mathbf{X}} = 0$ and $\mathbf{J}$ is the Fisher information for $\psi(\phi)$ in the model (3.6):

$$\mathbf{J} = \begin{bmatrix} \frac{1}{n}\mathbb{X}^T \left(\mathbf{V}^{-1} \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\right) \mathbb{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{E}_r^T \left(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}\right) \mathbf{E}_r \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{V}^{-1} \otimes \left(\frac{\mathbf{X}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n}\right) & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{E}_r^T \left(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}\right) \mathbf{E}_r \end{bmatrix}.$$

Then

$$\sqrt{n}(\hat{\phi} - \phi) \to N(\mathbf{0}, \mathbf{\Lambda}_0) \tag{S5.1}$$

where $\mathbf{\Lambda}_0 = \Psi(\Psi^T \mathbf{\Lambda} \Psi)^{\dagger} \Psi$, $\mathbf{\Lambda} = \mathbf{J}^{-1}$ is the asymptotic variance of the MLE under the full model, and Ψ is as follows:

$$\begin{bmatrix} \mathbf{K}_{rp}(\mathbf{I}_p \otimes \mathbf{\Gamma}_1) & \mathbf{K}_{rp}(\boldsymbol{\eta}^T \otimes \mathbf{I}_r) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{C}_r(\mathbf{\Gamma}_1\mathbf{\Omega}_1 \otimes \mathbf{I}_r - \mathbf{\Gamma}_1 \otimes \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T) & \mathbf{C}_r(\mathbf{\Gamma}_1 \otimes \mathbf{\Gamma}_1)\mathbf{E}_u & \mathbf{C}_r(\mathbf{\Gamma}_0 \otimes \mathbf{\Gamma}_0)\mathbf{E}_{r-u} \end{bmatrix}.$$

Furthermore, $\mathbf{\Lambda}^{-\frac{1}{2}}(\mathbf{\Lambda} - \mathbf{\Lambda}_0)\mathbf{\Lambda}^{-\frac{1}{2}} \geq 0$, so the spatial envelope model decreases the asymptotic variance.

Proof: We can rewrite $\boldsymbol{\beta}^*$ as follows

$$\begin{aligned}\boldsymbol{\beta}^* &= vec(\boldsymbol{\eta}^T\boldsymbol{\Gamma}_1^T) \\ &= \mathbf{K}_{rp}vec(\boldsymbol{\Gamma}_1\boldsymbol{\eta}) \\ &= \mathbf{K}_{rp}(\mathbf{I}_p \otimes \boldsymbol{\Gamma}_1)vec(\boldsymbol{\eta}) \\ &= \mathbf{K}_{rp}(\boldsymbol{\eta}^T \otimes \mathbf{I}_r)vec(\boldsymbol{\Gamma}_1).\end{aligned} \tag{S5.2}$$

Therefore, the derivatives of ψ_1 with respect to ϕ_1^T is

$$\frac{\partial\psi_1}{\partial\phi_1^T} = \frac{\partial\boldsymbol{\beta}^*}{\partial(vec(\boldsymbol{\eta}))^T} = \frac{\partial\left[\mathbf{K}_{rp}(\mathbf{I}_p \otimes \boldsymbol{\Gamma}_1)vec(\boldsymbol{\eta})\right]}{\partial(vec(\boldsymbol{\eta}))^T} = \mathbf{K}_{rp}(\mathbf{I}_p \otimes \boldsymbol{\Gamma}_1),$$

and the derivatives of ψ_1 with respect to ϕ_2^T is

$$\frac{\partial\psi_1}{\partial\phi_2^T} = \frac{\partial\boldsymbol{\beta}^*}{\partial(vec(\boldsymbol{\Gamma}))^T} = \frac{\partial\left[\mathbf{K}_{rp}(\boldsymbol{\eta}^T \otimes \mathbf{I}_r)vec(\boldsymbol{\Gamma}_1)\right]}{\partial(vec(\boldsymbol{\Gamma}_1))^T} = \mathbf{K}_{rp}(\boldsymbol{\eta}^T \otimes \mathbf{I}_r). \tag{S5.3}$$

It is clear that $\frac{\partial\psi_1}{\partial\phi_3^T} = \frac{\partial\psi_1}{\partial\phi_4^T} = \mathbf{0}$.

The derivative of $\frac{\partial\psi_2}{\partial\phi_1^T}$ to $\frac{\partial\psi_2}{\partial\phi_4^T}$ are similar to those in Cook, Li, and Chiaromonte (2010). Having these derivatives together lead to obtain (4.3).

The asymptotic distribution (S5.1) follows from Shapiro (1986). In order to prove that $\boldsymbol{\Lambda}_0 \leq \boldsymbol{\Lambda}$, we have

$$\boldsymbol{\Lambda}_0 - \boldsymbol{\Lambda} = \mathbf{J}^{-1} - \Psi(\Psi^T\boldsymbol{\Lambda}\Psi)^{\dagger}\Psi = \mathbf{J}^{-\frac{1}{2}}\left[\mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\Psi(\Psi^T\boldsymbol{\Lambda}\Psi)^{\dagger}\Psi\mathbf{J}^{\frac{1}{2}}\right]\mathbf{J}^{-\frac{1}{2}}$$

Since the matrix $\mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\Psi(\Psi^T\boldsymbol{\Lambda}\Psi)^{\dagger}\Psi\mathbf{J}^{\frac{1}{2}}$ is the projection on to orthogonal complement of $span(\mathbf{J}^{\frac{1}{2}}\Psi)$, it is positive semidefinite, which implies that $\boldsymbol{\Lambda}_0 - \boldsymbol{\Lambda}$ is also positive semidefinite. In addition, we have

$$\boldsymbol{\Lambda}^{-\frac{1}{2}}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\boldsymbol{\Lambda}^{-\frac{1}{2}} = \mathbf{I}_{pr+r(r+1)/2} - \mathbf{J}^{\frac{1}{2}}\Psi(\Psi^T\boldsymbol{\Lambda}\Psi)^{\dagger}\Psi\mathbf{J}^{\frac{1}{2}}$$

which proves the last statement of the theorem.

S6 Proof of Corollary 1

In this section, we restate and proof the corollary 1.

Corollary 1: The asymptotic variance (avar) of $\sqrt{n}\boldsymbol{\beta}^*$ can be written as

$$avar(\sqrt{n}\boldsymbol{\beta}^*) = \mathbf{K}_{rp}\left\{\left(\frac{\mathbf{X}^T\boldsymbol{\rho}(\boldsymbol{\theta})^{-1}\mathbf{X}}{n}\right)^{-1} \otimes \boldsymbol{\Gamma}_1\boldsymbol{\Omega}_1\boldsymbol{\Gamma}_1^T + (\boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}_0)(\Psi_2^T\mathbf{J}\Psi_2)^\dagger(\boldsymbol{\eta} \otimes \boldsymbol{\Gamma}_0^T)\right\}\mathbf{K}_{rp}^T \tag{S6.1}$$

where $\Psi_2 = \left(\frac{\partial\psi_1}{\partial\phi_2^T}, \frac{\partial\psi_2}{\partial\phi_2^T}\right)^T$.

Proof: Using lemma 1 and theorem 1, the asymptotic variance of $\sqrt{n}\boldsymbol{\beta}^*$ can be written as

$$avar(\sqrt{n}\boldsymbol{\beta}^*) = K_1(\Psi_1^T\mathbf{J}\Psi_1)^\dagger K_1^T + K_2(\Psi_2^T\mathbf{J}\Psi_2)^\dagger K_2^T$$

where $\Psi_1 = \left(\frac{\partial\psi_1}{\partial\phi_1^T}, \frac{\partial\psi_2}{\partial\phi_1^T}\right)^T$, $K_1 = \mathbf{K}_{rp}(\mathbf{I}_p \otimes \boldsymbol{\Gamma}_1)$ and $K_2 = \mathbf{K}_{rp}(\boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}_0)$. Using straightforward matrix multiplication and corollary D1 to D3 in Cook, Li, and Chiaromonte (2010) complete the proof.

S7 Proof of the comparison between the variance of the envelope and spatial envelope models

In this section, we restate and proof the equation (4.8).

For the simplify version of the spatial envelope and envelope, it can be shown that

$$\mathbf{V}_{SPEN}^{-\frac{1}{2}}\mathbf{V}_{EN}\mathbf{V}_{SPEN}^{-\frac{1}{2}} = \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\mathbf{I}_r + \left(\frac{(\sigma_0^2-\sigma_1^2)^2\left(1-\frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\right)}{(\sigma_0^2-\sigma_1^2)^2+\sigma_1^2\sigma_{\mathbf{X}}^2||\boldsymbol{\beta}||^2}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T, \tag{S7.1}$$

where $\mathbf{V}_{SPEN}$ shows the asymptotic variance of the spatial envelope model, $\mathbf{V}_{EN}$ shows the asymptotic variance of the envelope model, and $\sigma_{\mathbf{X}}^2$ denotes the variance of the variance of the $\mathbf{X}$ which is a $n\times 1$ vector.

Proof: For the simplified version of the mode, the asymptotic variance for two models are:

$$\begin{aligned} var(\sqrt{n}\boldsymbol{\beta}_{Env}) &= \frac{\sigma_1^2}{\sigma_{\mathbf{X}}^2}\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T + \frac{\sigma_0^2\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}}{\sigma_{\mathbf{X}}^2\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}+(\sigma_0^2-\sigma_1^2)^2}\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T,\\ var(\sqrt{n}\boldsymbol{\beta}^*) &= \frac{n\sigma_1^2}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T + \frac{n\sigma_0^2\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}+n(\sigma_0^2-\sigma_1^2)^2}\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T, \end{aligned}$$

therefore, to compare the variance of two models, we have

$$\begin{aligned} \mathbf{V}_{SPEN}^{-\frac{1}{2}}\mathbf{V}_{EN}\mathbf{V}_{SPEN}^{-\frac{1}{2}} &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T + \frac{n(\sigma_0^2-\sigma_1^2)^2+\sigma_1^2\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\eta}^T\boldsymbol{\eta}}{n(\sigma_0^2-\sigma_1^2)^2+n\sigma_1^2\sigma_{\mathbf{X}}^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T + \frac{n(\sigma_0^2-\sigma_1^2)^2+\sigma_1^2\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\eta}^T\boldsymbol{\eta}}{n(\sigma_0^2-\sigma_1^2)^2+n\sigma_1^2\sigma_{\mathbf{X}}^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &\pm \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\mathbf{I}_r + \left(-\frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2} + \frac{n(\sigma_0^2-\sigma_1^2)^2+\sigma_1^2\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\eta}^T\boldsymbol{\eta}}{n(\sigma_0^2-\sigma_1^2)^2+n\sigma_1^2\sigma_{\mathbf{X}}^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\mathbf{I}_r + \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\left(-1+\frac{\frac{n(\sigma_0^2-\sigma_1^2)^2}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}+\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}}{\frac{(\sigma_0^2-\sigma_1^2)^2}{\sigma_{\mathbf{X}}^2}+\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\mathbf{I}_r + \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\left(-1+1+\frac{(\sigma_0^2-\sigma_1^2)^2\left(\frac{n}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}-\frac{1}{\sigma_{\mathbf{X}}^2}\right)}{\frac{(\sigma_0^2-\sigma_1^2)^2}{\sigma_{\mathbf{X}}^2}+\sigma_1^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T\\ &= \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\mathbf{I}_r + \frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}\left(\frac{(\sigma_0^2-\sigma_1^2)^2\left(\frac{n\sigma_{\mathbf{X}}^2}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}-1\right)}{(\sigma_0^2-\sigma_1^2)^2+\sigma_1^2\sigma_{\mathbf{X}}^2\boldsymbol{\eta}^T\boldsymbol{\eta}}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T \end{aligned}$$

Since $\boldsymbol{\eta}^T\boldsymbol{\eta} = ||\boldsymbol{\eta}||^2 = ||\boldsymbol{\beta}||^2$, therefore we have

$$\frac{\mathbf{V}_{SPEN}^{-\frac{1}{2}}\mathbf{V}_{EN}\mathbf{V}_{SPEN}^{-\frac{1}{2}}}{\frac{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}}{n\sigma_{\mathbf{X}}^2}} = \mathbf{I}_r + \left(\frac{(\sigma_0^2 - \sigma_1^2)^2\left(\frac{n\sigma_{\mathbf{X}}^2}{\mathbf{X}^T\boldsymbol{\rho}^{-1}(\boldsymbol{\theta})\mathbf{X}} - 1\right)}{(\sigma_0^2 - \sigma_1^2)^2 + \sigma_1^2\sigma_{\mathbf{X}}^2||\boldsymbol{\beta}||^2}\right)\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T$$

S8 Preliminary Analysis for the Real Data

In this section, we provide the estimated Moran's autocorrelation coefficient (also called Moran's I) and empirical variogram for the real data. Moran's I is an extension of the Pearson correlation and measures spatial autocorrelation in the data (Cliff and Ord, 1973). For a a vector of data s, Moran's I is

$$MI = \frac{n}{S_0}\frac{\sum_{i=1}^n\sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

where $\bar{x}$ denotes the mean of the observation, w_{ij} is the weight between observation i and j, and S_0 is the sum of all weights i.e. $S_0 = \sum_{i=1}^n\sum_{j=1}^n w_{ij}$. The weights w_{ij}, are chosen to be the inverse of the distance between observation i and j. Using Moran's I, one can test the existence of the spatial autocorrelation where the null hypothesis is that there is no correlation versus the alternative hypothesis of there exists the spatial statistics. Table 4 presents the results of Moran's I for all the variables in the study. Based on these results, we can reject the null hypothesis that there is zero spatial autocorrelation present in the data for each variable.

Table 1: Moran's I for different variables in the study.

Variable	observed	expected	sd	p.value
Ozone	0.4498559	-0.003731343	0.02014298	0
Carbon monoxide	0.08161912	-0.003731343	0.01918668	8.650319e-06
Sulfur dioxide	0.2425074	-0.003731343	0.01981788	0
Lead	0.234758	-0.003731343	0.01924146	0
Nitrogen dioxide	0.4414368	-0.003731343	0.02013472	0
Nitrogen monoxide	0.1665705	-0.003731343	0.01911524	0
PM 2.5	0.2449143	-0.003731343	0.02014268	0
PM 10	0.4063382	-0.003731343	0.01967082	0

In addition, to test the existence of the spatial correlation in the data, one common approach is to look at the patterns of the empirical variograms for the data in the preliminary analysis. We used the Matern covariance function for the real data analysis. Using this covariance function makes the computation faster and it is one of the most common covariance function used in analyzing the air pollution data. Figure 1 shows the empirical variogram of the responses. These plots show that using a Matern covariance function is reasonable.

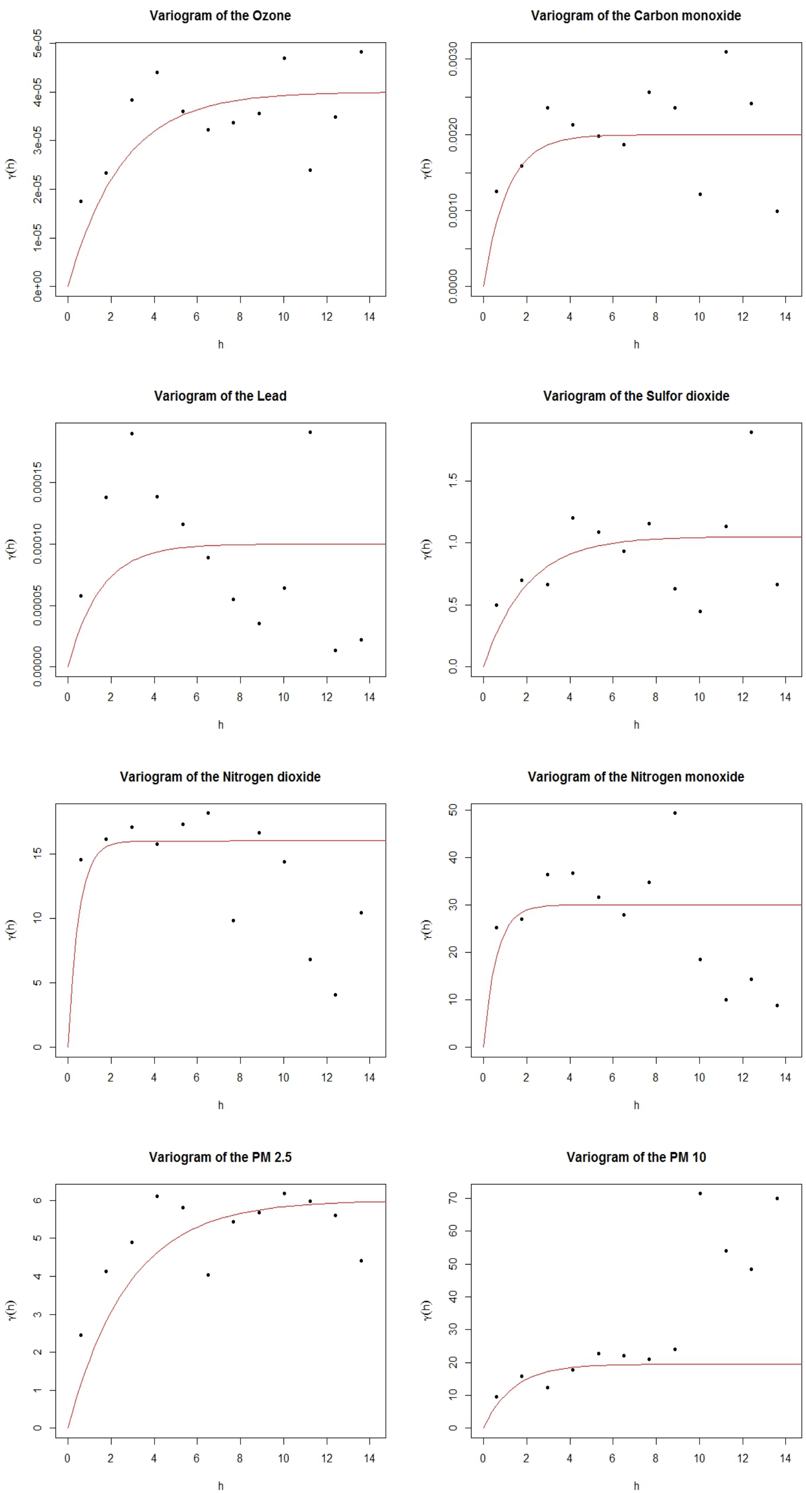

Figure 1: The empirical variogram of different responses in our study. These plots shows that using a Matern covariance function is reasonable.

S9 Estimated Regression Coefficients

In this section, we provide the estimated regression coefficients and their standard deviation for traditional envelope model and our proposed model. As it can be seen the standard deviation for the estimated coefficients based on our proposed model is smaller than those calculated by traditional envelope model.

Table 2: Regression Coefficients (asymptotic standard deviation) using envelope the air pollution data in northeastern United States.

Variable	Relative humidity	Temperature	Wind
Ozone	0.068 (0.388)	-0.083 (0.493)	-0.034 (0.303)
Carbon monoxide	-0.008 (0.051)	0.014 (0.064)	0.004 (0.040)
Lead	-0.016 (0.094)	0.022 (0.120)	0.008 (0.074)
Nitrogen dioxide	-0.050 (0.515)	0.148 (0.564)	0.037 (0.406)
Nitrogen monoxide	-0.032 (0.442)	0.157 (0.553)	0.001 (0.346)
Sulfur dioxide	-0.029 (0.381)	0.196 (0.487)	0.007 (0.297)
PM10	0.013 (0.353)	0.188 (0.440)	-0.021 (0.276)
PM2.5	0.033 (0.343)	-0.162 (0.581)	-0.011 (0.261)

Table 3: Regression coefficients (asymptotic standard deviation) using spatial envelope the air pollution data in northeastern United States.

Variable	Relative humidity	Temperature	Wind
Ozone	0.007 (0.178)	-0.004 (0.083)	-0.004 (0.033)
Carbon monoxide	0.011 (0.005)	0.014 (0.064)	-0.001 (0.001)
Lead	-0.001 (0.014)	0.002 (0.120)	0.001 (0.004)
Nitrogen dioxide	0.072 (0.021)	0.348 (0.121)	-0.037 (0.046)
Nitrogen monoxide	0.062 (0.022)	0.457 (0.115)	-0.084 (0.023)
Sulfur dioxide	-0.613 (0.111)	0.196 (0.006)	0.004 (0.096)
PM10	-0.013 (0.025)	0.188 (0.024)	-0.098 (0.026)
PM2.5	0.116 (0.143)	0.162 (0.051)	0.003 (0.016)

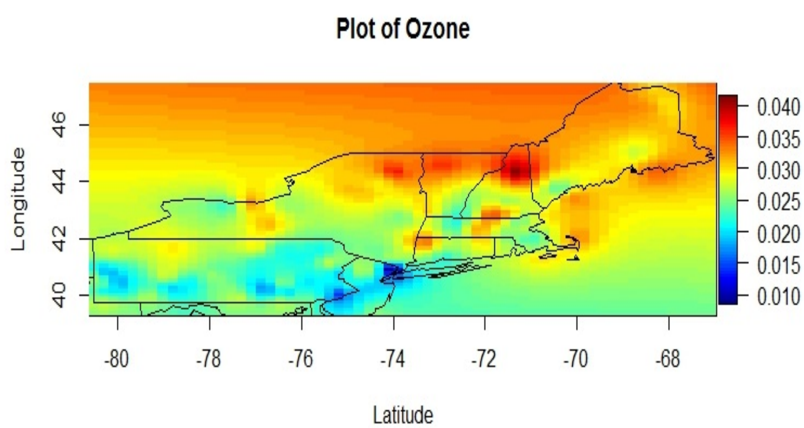

Figure 2: Prediction plot for the log of the ground level Ozone for the study area. Ozone level is not high in the study area. The north part of New Hampshire seems to have the highest value for the Ozone.

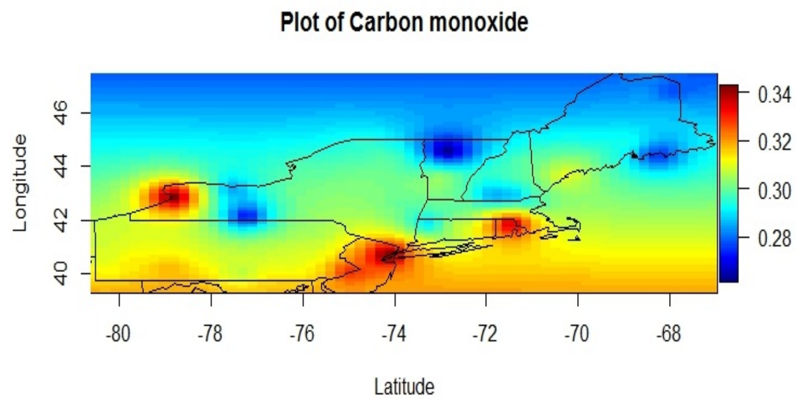

Figure 3: Prediction plot for carbon monoxide (CO) for the study area. Carbon monoxide is moderately low in the study area. CO is high in Rhode Island, New York, New Jersey, and Buffalo which are highly populated and therefore there will be a lots of car and usage of fossil fuels which leads to high concentration of carbon monoxide in the air.

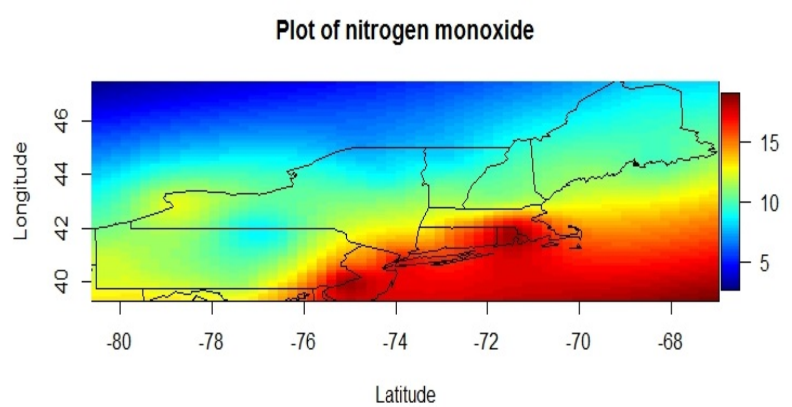

Figure 4: Prediction plot for the Nitrogen monoxide for the study area. Nitrogen monoxide is high in New York and New Jersey and moderately high almost every place in the study area.

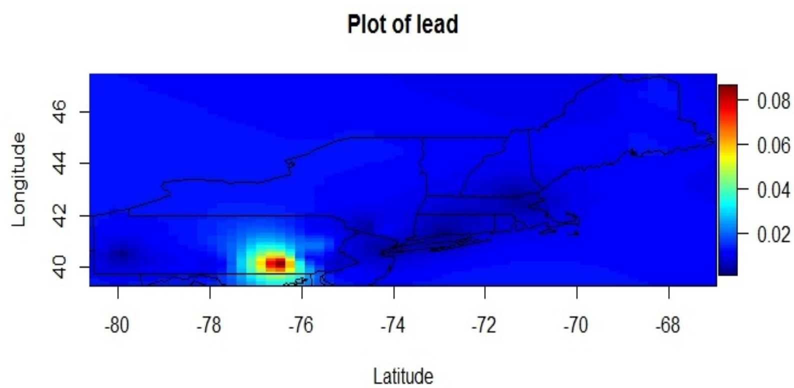

Figure 5: Prediction plot for lead for the study area. Lead is high in Harrisburg and Lancaster.

S10 Prediction Plot for Response Variables

Bibliography

Cliff, A. D., and Ord, J. K. (1973). *Spatial Autocorrelation.* Pion, London.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, pp. 927–960.

Goulard, M and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* 24 (3), pp. 269–286.

Seber, G. A. F. (2008) *A matrix handbook for statisticians.* Volume 15. John Wiley & Sons.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of American Statistical Association* 81, pp. 142–149.

Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models, *Environmetrics*, 18 (2), pp. 125–139.