New Parsimonious Multivariate Spatial Model: Spatial Envelope

Hossein Moradi Rekabdarkolae, Qin Wang, Zahra Naji, Montserrat Fuentes

a Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284, USA

October 24, 2017

Abstract

Dimension reduction provides a useful tool for analyzing high dimensional data. The recently developed Envelope method is a parsimonious version of the classical multivariate regression model through identifying a minimal reducing subspace of the responses. However, existing envelope methods assume an independent error structure in the model. While the assumption of independence is convenient, it does not address the additional complications associated with spatial or temporal correlations in the data. In this article, we introduce a Spatial Envelope method for dimension reduction in the presence of dependencies across space. We study the asymptotic properties of the proposed estimators and show that the asymptotic variance of the estimated regression coefficients under the spatial envelope model is smaller than that from the traditional maximum likelihood estimation. Furthermore, we present a computationally efficient approach for inference. The efficacy of the new approach is investigated through simulation studies and an analysis of an Air Quality Standard (AQS) dataset from the Environmental Protection Agency (EPA).

Keyword: Dimension reduction, Grassmanian manifold, Matern covariance function, Spatial dependency.

1 Introduction

In many research areas, such as health science (Lave, and Seskin, 1973; Liang, Zeger, and Qaqish, 1992), environmental sciences (Guinness et al., 2014), and...
business (Cooper, Schindler, and Sun, 2003), etc., it is common to observe multiple outcomes simultaneously. The traditional multivariate linear model has proved to be useful in these cases to understand the relationship between response variables and predictors. Mathematically, the model is typically presented as:

\[ Y = \alpha + \beta X + \epsilon, \]  

where \( Y \in \mathbb{R}^r \) denotes the response vector, \( X \in \mathbb{R}^p \) is a vector predictor, \( \alpha \in \mathbb{R}^r \) denotes the vector of intercept, \( \beta \in \mathbb{R}^{(r \times p)} \) is the matrix of regression coefficients, and \( \epsilon \sim N_r(0, \Sigma) \) is an error vector with \( \Sigma \geq 0 \) being an unknown covariance matrix (Christensen, 2001). In order to completely specify a multivariate linear model, there are \( r \) unknown intercepts, \( p \times r \) unknown parameters for the matrix of regression coefficients, and \( r(r + 1)/2 \) unknown parameters to specify an unstructured covariance matrix. Therefore, one must estimate \( r + pr + r(r + 1)/2 \) parameters which can be large with the increase of either \( r \) or \( p \) or both.

Based on the observation that some linear combinations of \( Y \) do not depend on any of the predictors in some cases, Cook, Li, and Chiaromonte (2010) proposed the Envelope method as a parsimonious version of the classical multivariate linear model. This approach separates the \( Y \) into material and immaterial parts, thereby allowing gains in estimation efficiency compared to the usual maximum likelihood estimation. The envelope approach constructs a link between the mean function and covariance matrix using a minimal reducing subspace such that the resulting number of parameters will be maximally reduced. Cook, Li, and Chiaromonte (2010) showed that the envelope estimator has asymptotically less variation compared to the standard maximum likelihood estimator (MLE). Along the same line, the idea of envelope has been further developed from both theoretical and computational points of view in a series of papers by Cook, Helland, and Su (2013); Cook and Su (2013); Cook, Su, and Yang (2015); Cook, Forzani, and Zhang (2015); Cook and Zhang (2015, 2016); Cook, Su, and Yang (2016); Su and Cook (2011, 2012, 2013); Guo et al. (2016), Li and Zhang (2017), Zhang and Li (2017), Park, Su, and Zhu (2017), and Khare, Pal, and Su (2017).

Existing envelope methodology assumes observations are taken under identical conditions where independence is assured. While models based on the independence assumption are extremely useful, their use is limited in applications where the data has inherent dependency (Cressie, 1993). For example, in environment monitoring, each station collects data concerning several pollutants such as ozone, carbon monoxide, nitrogen dioxide, etc. These data have a special type of dependency which is called spatial correlation. In this paper, we introduce a Spatial Envelope approach for spatially correlated data. This new approach addresses the impact of spatial correlation among observations in the
model and thus provides more efficient estimators than the traditional multivariate linear model and linear coregionalization model (Zhang, 2007). Accounting for the intrinsic spatial correlation allows the appropriate inference on aforementioned data.

The rest of the paper is organized as follows: in section 2, we briefly review envelope methodology. The spatial envelope is detailed in Section 3. Section 4 and 5 provide asymptotic variance and prediction properties of the proposed method. Section 6 and 7 contain a simulation study and the analysis of the northeastern United State air pollution data. We conclude the article with a short discussion in Section 8. All technical details are provided in the Appendix.

2 Brief Review of envelope

For model (1), suppose that we can find an orthogonal matrix \((\Gamma_1, \Gamma_0) \in \mathbb{R}^{r \times r}\) that satisfies the following two conditions: (i) \(\text{span}(\beta) \subseteq \text{span}(\Gamma_1)\), and (ii) \(\Gamma_1^T Y\) is conditionally independent of \(\Gamma_0^T Y\) given \(X\). That is, \(\Gamma_0^T Y\) is marginally independent of \(X\) and conditionally independent of \(X\) given \(\Gamma_1^T Y\). Then, we can rewrite \(\Sigma\) as

\[
\Sigma = P_{\Gamma_1} \Sigma P_{\Gamma_1} + Q_{\Gamma_1} \Sigma Q_{\Gamma_1}, \tag{2}
\]

where \(P_{(\cdot)}\) represents an orthogonal projection operator with respect to the standard inner product and \(Q_{(\cdot)} = I - P_{(\cdot)}\) is the projection onto its complement space. [Cook, Li, and Chiaromonte (2010)] used this idea to construct the unique smallest subspace \(\text{span}(\Gamma_1)\) that satisfies (2) and contains \(\text{span}(\beta)\). In summary, the goal is to find a subspace \(\Gamma_1 \subseteq \mathbb{R}^r\) such that

\[
Q_{\Gamma_1} Y | X \sim Q_{\Gamma_1} Y, \tag{3a}
\]

\[
Q_{\Gamma_1} Y \perp P_{\Gamma_1} Y | X. \tag{3b}
\]

where \(\perp\) means statistical independence. This minimal subspace is called the \(\Sigma\)-envelope of \(\text{span}(\beta)\) in full and the envelope for brevity. \(\Gamma_1^T Y\) and \(\Gamma_0^T Y\) are referred as material and immaterial parts of \(Y\), respectively, where \(u \leq r\), is referred as the dimension of the envelope subspace.

Following the envelope idea, model (1) can be rewritten as

\[
Y = \alpha + \Gamma_1 \eta X + \epsilon, \tag{4}
\]

where \(\beta = \Gamma_1 \eta, \eta \in \mathbb{R}^{u \times p}\), and \(\Sigma = \Sigma_0 + \Sigma_1\) such that \(\Sigma_0 = Q_{\Gamma_1} \Sigma Q_{\Gamma_1}\) being the variance of the immaterial part of response and \(\Sigma_1 = P_{\Gamma_1} \Sigma P_{\Gamma_1}\) being the variance of the material part of response. [Cook, Li, and Chiaromonte (2010)] showed that \(\Sigma = \Gamma_1 \Omega_1 \Gamma_1^T + \Gamma_0 \Omega_0 \Gamma_0^T\) where \(\Omega_1 = \text{var}(\Gamma_1^T Y) \in \mathbb{R}^{u \times u}\) and \(\Omega_0 = \).
\[
\text{var}(\Gamma_0^T \mathbf{Y}) \in \mathbb{R}^{(r-u) \times (r-u)} \text{ are unknown positive definite matrices with } 0 < u \leq r.
\]
Here, one only needs to estimate \( r + pu + r(r+1)/2 \) parameters. The difference in the number of parameters between the envelope and classical multivariate regression is \( p(r-u) \). More details can be found in Cook, Li, and Chiaromonte (2010) and the references therein.

3 New Spatial Envelope

In this section, we detail the spatial envelope method. We start with a review of spatial multivariate model, then drive the likelihood function of spatial envelope model, and show the computational steps for the parameter estimation. Let \( \mathbf{Y}(s_i) = (Y_1(s_i), \ldots, Y_r(s_i)) \) be an \( r \)-variate stochastic spatial response vector along with \( p \) regressors \( (X_1(s_i), \ldots, X_p(s_i)) \) observed at locations \( s = \{s_1, s_2, \ldots, s_n; \ s_i \in \mathbb{R}^2; i = 1, 2, \ldots, n\} \). The multivariate spatial regression model can be written as:

\[
\mathbf{Y}(s) = \alpha^T \otimes \mathbf{1}_n + \mathbf{X}(s)\beta^T + \mathbf{\epsilon}(s), \quad (5)
\]

where \( \mathbf{Y}(s) \) denotes the \( n \times r \) response matrix, \( \mathbf{X}(s) \) is the \( n \times p \) matrix of covariates, and \( \otimes \) denotes the Kronecker product. Furthermore, \( \alpha \) denotes the \( r \times 1 \) vector of intercept, \( \mathbf{1}_n \) is an \( n \times 1 \) column vector with 1 at each entry, \( \beta \) is the \( r \times p \) matrix of regression coefficients, and \( \mathbf{\epsilon} \) is a multivariate spatial process with mean 0. We assume that the data generating process is second order stationary and the covariance of the response vectors \( \mathbf{Y}(s_i) \) and \( \mathbf{Y}(s_j) \) at two sites \( s_i \) and \( s_j \) is a function of distance between the two sites. Namely the covariance can be written as:

\[
\text{Cov}(\mathbf{Y}(s_i), \mathbf{Y}(s_j)) = C_{ij}(h), \quad h = ||s_i - s_j||, \quad (6)
\]

where \( || \cdot || \) denotes Euclidean distance. The function \( \text{C}(h) = \{C_{ij}(h)\} \) is the multivariate covariogram, \( C_{ij}(\cdot) \) is the direct covariogram for \( i = j \) and cross-covariogram for \( i \neq j \). By adopting the proportional correlation model (Chiles and Delfiner, 1999), the spatial covariance function can be written as

\[
\text{C}(h) = \mathbf{V}\rho(h), \quad (7)
\]

where \( \mathbf{V} \) is an \( r \times r \) positive definite matrix and \( \rho(h) \) is a positive semidefinite correlation matrix (Wackernagel, 2003). Estimating the correlation function solely from the data without any structural assumptions is difficult and sometimes infeasible. Usually, it is assumed that the form of the correlation function is a known function but with unknown parameters \( \theta \), which control range, smoothness, and other characteristics of the correlation function. Thus instead of \( \rho(h) \),
we use $\rho(h, \theta)$ to represent unknown parameters $\theta$ in the correlation function. For simplicity of notation, $\rho(h, \theta)$ is denoted by $\rho(\theta)$ throughout the rest of the paper.

To illustrate the estimation, we use a $\text{vec}$ operator on the response matrix. That is, let $Y(s) = \text{vec}(Y(s))$ be an $nr \times 1$ vector for the vectorized response variable, and $X(s) = I_r \otimes X(s)$ be an $nr \times pr$ block diagonal matrix having $X_i(s)$ as blocks. Thus, the vectorized version of the multivariate spatial linear model can be written as:

$$Y(s) = \alpha \otimes 1_n + X(s)\beta^* + \epsilon^*(s). \quad (8)$$

where $\alpha$ is an $r \times 1$ vector of intercept, $\beta^* = \text{vec}(\beta^T)$ shows an $pr \times 1$ vector of regression coefficients, and $\epsilon^*(s)$ is an $nr \times 1$ vector of spatial errors with mean 0. With the use of proportional covariance model and the vectorization of the response matrix, the $nr \times nr$ covariance matrix of the response variables $\Sigma_Y$, can be written as $V \otimes \rho(\theta)$.

The likelihood function of model (8) is:

$$L(\alpha, \beta^*, V, \theta) = \left[\det(V \otimes \rho(\theta))\right]^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(Y(s) - \alpha \otimes 1_n - X(s)\beta^*)^T(V \otimes \rho(\theta))^{-1}(Y(s) - \alpha \otimes 1_n - X(s)\beta^*)\right\}, \quad (9)$$

where $\det(\cdot)$ denotes the determinant of the matrix. Suppose the response vector can be decomposed into the material and immaterial part, $Y_1$ and $Y_0$, respectively. From the envelope idea, $V$ can be written as $V_0 + V_1$ where $V_0 = Q\Gamma_1 V Q\Gamma_1$ denotes the covariance matrix associated with the immaterial part of response and $V_1 = P\Gamma_1 V P\Gamma_1$ denotes the covariance matrix associated with the material part where $\Gamma_1$ is the semi-orthogonal basis of $\text{span}(V_1)$. Hence, the covariance matrix of $Y$ can be written as follows:

$$\Sigma_Y = V \otimes \rho(\theta) = V_0 \otimes \rho(\theta) + V_1 \otimes \rho(\theta). \quad (10)$$

Let $0 < u \leq r$ denotes the structural dimension of the envelope, where $u$ can be selected using an information criterion such as AIC or BIC, or cross-validation. More details can be found in Cook, Li, and Chiaromonte (2010) and the references therein. Combining (9) and (10), we have

$$L^u(\alpha, \beta^*, V_0, V_1, \theta) = L^u_1(\alpha, \beta^*, V_1, \theta) \times L^u_2(\alpha, V_0, \theta), \quad (11)$$
with
\[ L_1^\prime(\alpha, \beta^*, V_1, \theta) = [\text{det}(V_1)]^{-\frac{r}{2}}[\text{det}(\rho(\theta))]^{-\frac{r}{2}} \times \exp\left\{ -\frac{1}{2}(\mathbf{Y}(s) - \alpha \otimes \mathbf{1}_n - \mathbf{X}(s)\beta^*)^T (V_1 \otimes \rho^{-1}(\theta)) (\mathbf{Y}(s) - \alpha \otimes \mathbf{1}_n - \mathbf{X}(s)\beta^*) \right\} , \]
\[ L_2^\prime(\alpha, V_0, \theta) = [\text{det}(V_0)]^{-\frac{r}{2}}[\text{det}(\rho(\theta))]^{-\frac{r}{2}} \times \exp\left\{ -\frac{1}{2}(\mathbf{Y}(s) - \alpha \otimes \mathbf{1}_n)^T (V_0 \otimes \rho^{-1}(\theta)) (\mathbf{Y}(s) - \alpha \otimes \mathbf{1}_n) \right\} , \]
\[ (12) \]

where $\dagger$ denotes the Moore-Penrose inverse and $\text{det}(A)$ denotes the product of non-zero eigenvalues of a non-zero symmetric matrix $A$. The likelihood in equation (9) can be factorized as equation (11) from $\text{span}(\beta) \subseteq \text{span}(V_1)$, and $(V_0 \otimes \rho(\theta))\beta^* = 0$. This factorization is detailed in the Appendix, section 9.1.

The objective is to maximize the likelihood in (11) over $\beta^*, V_0, V_1$, and $\theta$ subject to the constraints:
\[ \text{span}(\beta) \subseteq \text{span}(V_1), \]
\[ V_0V_1 = 0 \]
\[ (13) \]

As mentioned by Cook, Li, and Chiaromonte (2010), the gradient-based algorithms for Grassmann optimization (Edelman, Arias, and Smith [1998]) require a coordinate version of the objective function which must have continuous directional derivatives. The optimization depends on minimizing the logarithm of $\mathbf{D}$ over the Grassmann manifold $\mathbb{G}^{r \times u}$, where
\[ \mathbf{D} = \text{det}(P_{V_1} \Sigma_{\text{res}} P_{V_1} + Q_{V_1} \Sigma_{\mathbf{Y}} Q_{V_1}) , \]
and $\mathbf{D}$ is the partially maximized likelihood function. The derivation of $\mathbf{D}$ is detailed in the Appendix, section 9.2. Let $\Gamma_1$ be the semi-orthogonal basis for $\text{span}(V_1)$ and $\Gamma_0$ be the semi-orthogonal basis for $\text{span}(V_0)$. Then $\eta = \Gamma_1^T \beta$, $\Omega_1 = \Gamma_1^T \Sigma_{\text{res}} \Gamma_1$ and $\Omega_0 = \Gamma_0^T \Sigma_{\mathbf{Y}} \Gamma_0$, where $\Sigma_{\mathbf{Y}}$ and $\Sigma_{\text{res}}$ are the marginal covariance matrix of $\mathbf{Y}$ and the residual covariance matrix, respectively. Let $\log \text{det}(\cdot)$ denote the composite function $\log \circ \text{det}(\cdot)$. Then, the coordinate form of the log $\mathbf{D}$
\[ \log \mathbf{D} = \log \det\left( \Gamma_1^T \left( H^T \rho^{-1}(\theta) H - H^T \rho^{-1}(\theta) G \left(G^T \rho^{-1}(\theta) G\right)^{-1} G^T \rho^{-1}(\theta) H \right) \right) + \Gamma_0^T \left( H^T \rho^{-1}(\theta) H \right) \Gamma_0 \]
\[ (14) \]

where $H = \mathbf{Y} - \bar{\mathbf{Y}} \otimes \mathbf{1}_n$, and $G = \mathbf{X} - \bar{\mathbf{X}} \otimes \mathbf{1}_n$.

In order to obtain the parameters of spatial envelope model, the objective function (14) can be minimized by the gradient based Grassmann optimization. To do this, first obtain an initial value for $\Sigma_{\mathbf{Y}}$, $\Sigma_{\text{res}}$, and $\beta_{\text{MLE}}$, the marginal
covariance matrix of $\mathbb{Y}$, the residual covariance matrix, and the maximum likelihood estimate for $\beta$ from the fit of the full model (8). Set $\Theta^1 = \Theta^0$ where $\Theta = \{\theta, V_0, V_1\}$ and $V_0$ and $V_1$ can be obtained using traditional envelope model and $\theta$ can be obtained using linear coregionalization model. Then, we estimate $P_{\mathbb{Y}m}$ by minimizing the objective function (14) over the Grassmann manifold $G^{(r \times u)}$, and estimate $P_{\mathbb{Y}m}$ by $\bar{P}_{\mathbb{Y}m} = I - \hat{P}_{\mathbb{Y}m}$. In order to update the covariance function of material and immaterial parts of the spatial envelope, fix $\theta^m$ and estimate $V_0^m$ and $V_1^m$ by $\hat{V}_0^m = \bar{P}_{\mathbb{Y}m} \hat{\Sigma}_m^m \bar{P}_{\mathbb{Y}m}$ and $\hat{V}_1^m = \bar{P}_{\mathbb{Y}m} \hat{\Sigma}_m^m \bar{P}_{\mathbb{Y}m}$. Then, fix $\hat{V}_0^m$ and $\hat{V}_1^m$ and maximize $L^{(u)}(\alpha, \beta, V_0^m, V_1^m, \theta^m)$ over $\theta$ by solving the following minimization problem using numerical algorithm such as Newton-Raphson method:

$$\hat{\theta}^m = \underset{\theta}{\text{argmax}} \{ r \text{det}(\rho(\theta)) \}$$

$$\frac{1}{2} \text{tr} \left( \left( Q \left( \rho^{-\frac{1}{2}}(\theta; \bar{\phi}) \right) \rho^{-\frac{1}{2}}(\theta) H \right) V_0^{m\text{-}1} \left( Q \left( \rho^{-\frac{1}{2}}(\theta; \bar{\phi}) \right) \rho^{-\frac{1}{2}}(\theta) H \right)^T + \rho^{-\frac{1}{2}}(\theta) H V_0^{m\text{-}1} H^T \rho^{-\frac{1}{2}}(\theta) \right).$$

(15)

Now, update $\hat{\Sigma}_m^m$ and $\hat{\Sigma}_{\text{res}}^m$ using the new estimate for $V_0, V_1$, and $\theta$. Then, check the convergence. If $||\Theta_0^{m+1} - \Theta_0^m|| < \delta$ where $\delta$ is a pre-specified tolerance level, then stop the iteration, output the final spatial envelope estimators and estimate $\beta$ by $\hat{\beta} = \hat{P}_{V_1} \hat{\beta}_{MLE}$; otherwise, set $m := m + 1$ and redo the procedure. Finally, estimate the intercept by $\hat{\alpha} = \bar{Y} - \bar{X} \hat{\beta}^T$.

4 Theoretical Properties

In what follows, we study the asymptotic properties of the spatial envelope parameter estimates. The regression coefficients can be written as $\beta = \Gamma_1 \eta$. Furthermore, $V_0 = \Gamma_0 \Omega_0 \Gamma_0^T$ and $V_1 = \Gamma_1 \Omega_1 \Gamma_1^T$ are the covariance of the immaterial part and material part to the regression, respectively. Therefore, aside from the intercept, the parameters of spatial envelope model in equation (8) can be combined into the vector as follows:

$$\phi = \begin{bmatrix} \text{vec}(\eta) \\ \text{vec}(\Gamma_1) \\ \text{vech}(\Omega_1) \\ \text{vech}(\Omega_0) \end{bmatrix} \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix}$$

(16)

where the $\text{vec}(\cdot)$ denotes the vector operator and $\text{vech}(\cdot)$ denotes vector half operator. For background on these operators, see Seber (2008). Here we focus on the following parameters under the spatial envelope model:

$$\psi(\phi) = \begin{bmatrix} \beta^* \\ \text{vech}(V) \end{bmatrix} = \begin{bmatrix} \text{vec}(\eta^T \Gamma_1^T) \\ \text{vech} \left( (\Gamma_1 \Omega_1 \Gamma_1^T + \Gamma_0 \Omega_0 \Gamma_0^T) \right) \end{bmatrix} \equiv \begin{bmatrix} \psi_1(\phi) \\ \psi_2(\phi) \end{bmatrix}$$

(17)
Let
\[
\Psi = \begin{bmatrix}
\frac{\partial \psi_1}{\partial \phi_1} & \cdots & \frac{\partial \psi_1}{\partial \phi_r} \\
\frac{\partial \psi_2}{\partial \phi_1} & \cdots & \frac{\partial \psi_2}{\partial \phi_r}
\end{bmatrix}
\] (18)
denote the gradient matrix. Using this gradient matrix and following Cook, Li, and Chiaromonte (2010), we present the following asymptotic properties of proposed estimators.

**Lemma 1:** Suppose \( \bar{X} = 0 \), the Fisher information, \( J \), for \( \psi(\phi) \) in the model (8) is as follows:
\[
J = \begin{bmatrix}
V^{-1} \otimes (X^T \rho(\theta)^{-1}X) & 0 \\
0 & C_r (V^{-1} \otimes V^{-1}) E_r - \frac{1}{2} C_r \left[ \text{diag} (V^{-1} \otimes V^{-1}) \right] E_r
\end{bmatrix}
\] (19)
where \( C_r \in \mathbb{R}^{r(r+1)/2 \times r^2} \) is a contraction matrix which is defined such that for a given \( r \times r \) matrix \( A \), \( \text{vech}(A) = C_r \text{vec}(A) \), \( E_r \in \mathbb{R}^{r^2 \times r(r+1)/2} \) is an expansion matrix such that \( \text{vec}(A) = E_r \text{vech}(A) \), and \( \text{diag}(A) \) is the matrix with the diagonal elements of \( A \). The derivation of \( J \) is provided in the Appendix, section 9.3.

**Theorem 1:** Suppose \( \bar{X} = 0 \) and \( J \) is the Fisher information defined in lemma 1. Let \( \Lambda = J^{-1} \) be the asymptotic variance of the MLE under the full model. Then
\[
\sqrt{n}(\hat{\phi} - \phi) \to N(0, \Lambda_0)
\] (20)
where \( \Lambda_0 = \Psi(\Psi^T \Lambda \Psi)^{\dagger} \Psi \). Furthermore, \( \Lambda^{-\frac{1}{2}}(\Lambda - \Lambda_0)\Lambda^{-\frac{1}{2}} \geq 0 \), which means the asymptotic variance of the parameter estimation under the spatial envelope model is smaller than their estimate under MLE. Proof of this theorem can be found in the Appendix, section 9.4.

**Corollary 1:** The asymptotic variance (avar) of \( \sqrt{n}\beta^* \) can be written as
\[
\text{avar}(\sqrt{n}\beta^*) = X^T (V^{-1} \otimes \rho(\theta)^{-1})X \otimes \Gamma_1 \Omega_1 \Gamma_1^T + (\eta^T \otimes \Gamma_0)(\Psi_2^T J \Psi_2)^{\dagger} (\eta \otimes \Gamma_0^T)
\] (21)
where \( \Psi_2 = \begin{bmatrix} \frac{\partial \psi_1}{\partial \phi_2} & \frac{\partial \psi_2}{\partial \phi_2} \end{bmatrix}^T \). Proof of this corollary can be found in the Appendix, section 9.5.

5 Prediction

Prediction at an unsampled location is often a major objective of a spatial analysis. Let \( Y_{new} \) be the \( \text{vech}(X_{new}) \) of the new multivariate response at an unsampled location. The model then can be written as:
\[
\begin{pmatrix}
Y_{new} \\
\bar{Y}
\end{pmatrix} = \begin{pmatrix}
\alpha \otimes 1_{n_{new}} + X_{new} \beta^* \\
\alpha \otimes 1_n + X \beta^*
\end{pmatrix} + \begin{pmatrix}
\epsilon_{new} \\
\epsilon
\end{pmatrix} \sim N \left( \begin{pmatrix}
\alpha \otimes 1_N + X \beta^* \\
\alpha \otimes 1_n + X \beta^*
\end{pmatrix}, \Sigma \right).
\] (1)
where \( N = n + n_{new} \) and \( \Sigma \) is as follows

\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix} = \begin{pmatrix}
(V_0 + V_1) \otimes \rho_{new,new}(\theta) & (V_0 + V_1) \otimes \rho_{new,Y}(\theta) \\
(V_0 + V_1) \otimes \rho_{\chi,new}(\theta) & (V_0 + V_1) \otimes \rho_{\chi,Y}(\theta)
\end{pmatrix}.
\] (2)

The conditional distribution \( Y_{new} | Y \) is

\[
Y_{new} | Y, \alpha, \eta, V_0, V_1, \theta \sim N \left( \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (Y - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right),
\] (3)

where \( \mu_1 = \alpha \otimes 1_{new} + X_{new} \beta^* \) and \( \mu_2 = \alpha \otimes 1_n + X_\theta \beta^* \). Using the method described in section 3, one can estimate the parameters of the model and then from the conditional distribution (3) the \( E(Y_{new} | Y) \) can be estimated.

6 Simulation

In this section, we carry out a simulation study to evaluate the finite sample performance of the proposed spatial envelope model and to compare it with the traditional multivariate linear regression (MLR), linear coregionalization model (LCM; Zhang, 2007), and envelope (Cook, Li, and Chiaromonte, 2010). The data \{\( (X_1, Y_1), \ldots, (X_n, Y_n) \)\} are generated from the model

\[
Y = X_\theta \beta + \epsilon,
\] (4)

where \( Y_i \in \mathbb{R}^5, X_i \in \mathbb{R}^6 \), and the structural dimension \( u = 2 \). The matrix \((\Gamma_1; \Gamma_0)\) is obtained by orthogonalizing an \( 5 \times 5 \) matrix generated from uniform \((0, 1)\) variables. The elements in \( \eta \) follows a standard normal distribution, and \( \beta = \Gamma_1 \eta \). We generate \( \Sigma_{Y} = (\Gamma_1 \Omega_1 \Gamma_1^T + \Gamma_0 \Omega_0 \Gamma_0^T) \otimes \rho(\theta) \) where \( \Omega_1 = \{(0.9)^{|i-j|}\} \) and \( \Omega_0 = \{(-0.5)^{|i-j|}\} \). For the spatial correlation function \( \rho(\theta) \), we use the following Matern correlation function:

\[
\rho(h; \theta) = \frac{1}{2^{\theta_2 - 1} \Gamma(\theta_2)} \left( \frac{||h||}{\theta_1} \right)^{\theta_1} \kappa_{\theta_2} \left( \frac{||h||}{\theta_1} \right),
\]

where \( \theta = (\theta_1, \theta_2), \theta_1 > 0 \) is the range parameter, \( \theta_2 \) is the smoothness parameter, \( \Gamma(\cdot) \) is the Gamma function, and \( \kappa_{\theta_2} \) is the modified Bessel function of the second kind of order \( \theta_2 \) (Abramowitz and Stegun, 1964). Three error distributions of \( \epsilon \) are investigated. We assume \( \epsilon \) follows a normal distribution with mean 0 and covariance \( \Sigma \). For first error \( \Sigma = (\Gamma_1 \Omega_1 \Gamma_1^T + \Gamma_0 \Omega_0 \Gamma_0^T) \). This density serves as a benchmark where the errors are independent from each other. For the second scenario, let \( \epsilon \) follows a Matern covariance function with \( \theta_1 = 1 \) and \( \theta_2 = 0.5 \); This case represents a spatial correlation in the data with a short range of dependency. This case is an example of weak spatial correlation. Finally,
let $\epsilon$ follows a Matern covariance function with $\theta_1 = 5$ and $\theta_2 = 0.5$; This case represents a spatial correlation in the data with a long range of dependency. This case is an example of strong spatial correlation.

Sample size is 100, 225, and 400. There are two different ways to generate these samples. One is based on $10 \times 10$, $15 \times 15$ and $20 \times 20$ evenly spaced grids on $[0, 1]^2$, respectively. Another way is to randomly choose 100, 225, and 400 locations from a $101 \times 101$ grid on $[0, 1]^2$. We use both sampling procedures to check whether the spatial distribution of the observations has any impact on the proposed estimation. All results reported here are based on 200 replications from the simulation model in each scenario. In order to compare the different estimators, we use Leave One Out Cross-Validation (LOCV) method, which provides a convenient approximation for the prediction error under squared-error loss

$$MSPE = \frac{\sum_{i=1}^{n} (\hat{Y}(\cdot; i) - Y(s_{i, obs}))^2}{n},$$

where $Y(s_{i, obs})$ is the observe value for response in location $s_i$ and $\hat{Y}(\cdot; i) = Y(s_{i, obs})$ is the predicted values of $Y(s_i)$ computed with the $i$th row of the data removed. Tables 1 and 2 summarize the results of these simulations. These tables provide the LOCV for different methods and different error distributions.

Table 1: Prediction accuracy comparison based on the mean (standard deviation) of leave one out cross-validation (LOCV) for all 200 data sets from equally spaced samples. Smaller LOCV shows better performance.

| $\epsilon$ | n   | MLR     | LCM     | Envelope | Spatial Envelope |
|------------|-----|---------|---------|----------|-----------------|
| 1          | 100 | 19.02 (1.537) | 20.01 (1.754) | 13.71 (1.547) | 14.28 (1.644) |
|            | 225 | 18.49 (1.153) | 19.75 (1.659) | 11.49 (1.124) | 12.51 (1.234) |
|            | 400 | 18.27 (0.828) | 19.02 (1.002) | 10.37 (0.812) | 10.87 (0.989) |
| 2          | 100 | 102.79 (35.570) | 22.54 (3.246) | 91.98 (36.379) | 20.21 (1.988) |
|            | 225 | 101.57 (32.495) | 20.46 (2.897) | 89.24 (33.083) | 18.34 (1.450) |
|            | 400 | 99.98 (32.185) | 18.89 (2.051) | 88.95 (31.855) | 17.68 (1.056) |
| 3          | 100 | 117.79 (48.834) | 24.19 (4.125) | 119.08 (47.852) | 21.36 (2.353) |
|            | 225 | 103.22 (39.065) | 21.78 (3.278) | 104.73 (39.023) | 20.76 (2.012) |
|            | 400 | 99.08 (37.718) | 19.45 (3.001) | 100.39 (36.896) | 18.10 (1.651) |

From the summary of all three different error distributions, one can see that for the standard normal errors, where the observations are independent from each other, the spatial envelope provides comparable results to the envelope...
Table 2: Prediction accuracy comparison based on the mean (standard deviation) of leave one out cross-validation (LOCV) for all 200 data sets from random location samples. Smaller LOCV shows better performance.

| $\epsilon$ | n  | MLR      | LCM      | Envelope     | Spatial Envelope |
|------------|----|----------|----------|--------------|------------------|
| 1          | 100| 20.12 (1.613) | 21.01 (1.863) | 14.32 (1.699) | 14.98 (1.722)    |
|            | 225| 19.34 (1.231) | 19.68 (1.542) | 13.12 (1.234) | 13.19 (1.201)    |
|            | 400| 17.83 (0.804) | 18.22 (1.101) | 11.73 (0.718) | 12.37 (0.819)    |
| 2          | 100| 104.02 (36.702) | 23.32 (4.111) | 93.02 (30.433) | 19.21 (2.004)    |
|            | 225| 102.41 (34.521) | 21.41 (3.758) | 91.34 (27.211) | 17.34 (1.352)    |
|            | 400| 100.39 (30.822) | 19.20 (3.201) | 89.21 (25.581) | 16.68 (1.110)    |
| 3          | 100| 116.34 (45.089) | 25.21 (4.821) | 97.01 (43.021) | 20.79 (2.115)    |
|            | 225| 108.15 (34.211) | 22.35 (3.555) | 95.52 (31.774) | 18.92 (1.944)    |
|            | 400| 101.54 (32.102) | 20.44 (2.998) | 90.94 (30.234) | 17.03 (1.234)    |

method and both performs better than MLR and LCM. In error distributions 2 and 3 where there exists spatial dependency in the data, the spatial envelope method performed almost equally as well as they did in the cases without spatial dependency while original envelope loses its efficiency. In addition, spatial envelope outperformed LCM in both independent and dependent cases. Since spatial envelope takes the spatial correlation among observations into consideration, it provides more accurate results compared to the original envelope model. Furthermore, spatial envelope only uses the material part of the data which leads to a more efficient results compared to LCM which uses both material and immaterial part of the data. Therefore, we can conclude that the proposed spatial envelope model provided consistent estimates with good prediction accuracy in all error distributions considered. This result is consistent for both sampling methods which indicates the spatial distribution of the observations has minimal impact on the estimation.

As in Cook, Li, and Chiaromonte (2010), it is possible for an objective function defined on Grassmann manifolds to have multiple local optimal points. One way to check this is to run the simulation with different starting values and compare the results. In our numerical experiment, we have not find the local optima to be a problem for our method.

7 Application

In this section, we apply the proposed methodology to the air pollution data in the Northeastern United States. It is worth mentioning that the main purpose of
this data analysis is to provide an insight that how the proposed approach can be used to find the reduced response space in multivariate spatial data analysis. This data has drawn much attention from both statisticians and scientists in other areas. Researchers looked at this data from different points of view including, but not restricted to, climate change (Phelan et al., 2016), health science (Kioumourtzoglou et al., 2016), and air quality (Battye et al., 2016). These studies showed that relationships exist between air pollution and meteorological factors, such as wind, temperature and humidity. Most of the existing studies focus on one of these pollutants, but since correlation exists among these pollutants, it is beneficial to study them simultaneously.

The pollutants and weather data that we used in this study include the average levels of the following variables in January 2015. We choose a group of ambient air pollutants monitored by EPA because they present a high threat to human health. Specifically, we have 8 response variables: ground level ozone, sulfur dioxide ($SO_2$), carbon monoxide ($CO$), nitrogen dioxide ($NO_2$), nitrogen monoxide ($NO$), lead, PM 2.5, and PM 10. PM 10 includes particles less than or equal to 10 micrometers in diameter. Similarly, PM 2.5 includes particles less than or equal to 2.5 micrometers and is also called fine particle pollution. This data also includes the following meteorological variables: wind, temperature, and relative humidity as predictors. Along with this information, latitude and longitude of the monitoring locations are used to model the spatial structure in the data. Our study area consists of 9 states in the Northeast of the United States: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. This dataset is available at http://aqsdr1.epa.gov/aqstmp/airdata/download_files.html#. Cross-validation showed that the best choice for the structural dimension is 3. The Matern’s covariance parameters, $\theta_1$ and $\theta_2$, are estimated to be 0.51 and 0.91, respectively. This estimates shows the existence of spatial dependency in the data. The corresponding direction estimates ($\hat{\Gamma}_1$) from the spatial envelope are in Table 3.

By checking the estimated basis coefficients of the minimal subspace (directions), we can see Sulfur dioxide, Nitrogen dioxide, PM 10, and PM 2.5 dominate each of the three directions, respectively. Using fossil fuels creates sulfur dioxide, nitrogen monoxide, and nitrogen dioxide. The nitrogen monoxide will also become nitrogen dioxide in the atmosphere. Existence of the particles in the air leads to reduction in visibility and causes the air to become hazy when levels are elevated. Furthermore, since these particles can travel deeply into the human lungs, they can cause health problem such as lung cancer. The main source of these particles in the air is from pollutants emitted from power plants, industries
Table 3: The corresponding direction estimates using spatial envelope for the air pollution data in northeastern United States of America.

| Variable          | Direction 1 | Direction 2 | Direction 3 |
|-------------------|-------------|-------------|-------------|
| Ozone             | -0.0464     | 0.0432      | -0.0080     |
| Carbon monoxide   | 0.2840      | -0.3717     | -0.0179     |
| Lead              | -0.0739     | 0.0872      | 0.0008      |
| Nitrogen dioxide  | -0.5089     | 0.2612      | -0.4639     |
| Nitrogen monoxide | -0.3056     | -0.1137     | 0.2757      |
| Sulfur dioxide    | -0.5335     | 0.0241      | -0.2981     |
| PM10              | -0.3257     | -0.8667     | -0.0506     |
| PM2.5             | -0.4106     | 0.1394      | 0.7855      |
Figure 2: Prediction plot of the Sulfur dioxide for the study area. As it can be seen, the Sulfur dioxide is moderately high for the most part of the study area. Sulfur dioxide is extremely high in Johnstown where there exists a lot of defense manufacturing.

Figure 3: Prediction plot of the Nitrogen dioxide for the study area. The Nitrogen dioxide is high in Newark, New York, Philadelphia, and Rhodes Island which are all highly populated areas.
Figure 4: Prediction plot of the PM 10 for the study area. The PM 10 is high for most part of the study area especially in Philadelphia and Augusta.

Figure 5: Prediction plot of the PM 2.5 for the study area. The PM 2.5 is moderately high in almost every place in the study area especially in Pennsylvania state, Augusta, and middle of Vermont state.
and automobiles.

Figure 2 to 5 shows the prediction plots for the three pollutants with the largest impact. Figure 2 shows the prediction plot of the Sulfur dioxide for the study area. The Sulfur dioxide is moderately high for the most part of the study area. In addition, Sulfur dioxide is extremely high in Johnstown where there exists a lot of defense manufacturing. Figure 3 shows the prediction plot of the Nitrogen dioxide for the study area. The Nitrogen dioxide is high in Newark, New York, Philadelphia, and Rhodes Island which are all highly populated areas. Figure 4 shows the prediction plot of the PM 10 for the study area. The PM 10 is high for most part of the study area especially in Philadelphia and Augusta. Figure 5 shows the prediction plot of the PM 2.5 for the study area. The PM 2.5 is moderately high in almost every place in the study area especially in Pennsylvania state, Augusta, and middle of Vermont state. Prediction plots of the other variables can be found in Appendix, section 9.7.

The leave one out cross-validation for MLR, LCM, envelope, and spatial envelope are 7.537, 3.562, 4.876, and 1.978, respectively. The results of leave one out cross-validation show that spatial envelope outperforms other methods and provides more accurate prediction. In summary, we find out that the most important pollutants in January are particulates, sulfur, and nitrogen, and other pollutants have minimal effect. These statistical conclusions support the environmental chemical claim that in the cold weather, due to the fossil burning and inversion, sulfur dioxide, nitrogen dioxide, and particulate matters are the most important pollutants (Byers, 1959; Lægreid, Bockman, and Kaarstad, 1999).

8 Conclusion

Air pollution has a serious impact on human health. Research has greatly improved the understanding of each particular pollutant and their relationship with weather conditions. However, there are relatively few studies about the effects of meteorological variables on several pollutants together. Motivated by an analysis of air pollution in the northeastern United States, we proposed a new parsimonious multivariate spatial model. Emphasis of this work is placed on inference and constructing a method that can provide more efficient estimation for the parameters of interest than traditional maximum likelihood estimators through capturing the spatial structure in the data.

Our model is flexible enough to characterize complex dependency and cross-dependency structures of different pollutants. From a simulation study and real data analysis, we showed that the proposed spatial envelope model outperforms multivariate linear regression, envelope, and linear coregionalization models. This new approach provides more efficient estimation for regression coefficients.
compared to the traditional maximum likelihood approach.

The method presented in this paper is for a multivariate spatial response with separable covariance matrix. This framework can be also extended to the cases that the covariance matrix is non-separable. Another possible extension of current methodology is for the case with spatiotemporal responses. The investigation for these more general cases is under way.

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Volume 55. Courier Corporation.

Battye, W. H., Bray, C. D., Aneja, V. P., Tong, D., Lee, P., and Tang, Y. (2016). Evaluating ammonia (NH 3) predictions in the NOAA National Air Quality Forecast Capability (NAQFC) using in situ aircraft, ground-level, and satellite measurements from the DISCOVER-AQ Colorado campaign. *Atmospheric Environment* 140, pp. 342–351.

Byers, H. R. (1959). *General meteorology*. McGraw-Hill.

Chiles, J P and Delfiner, P. (1999). *Geostatistics: modeling spatial uncertainty.* Volume 497. John Wiley & Sons.

Christensen, R. (2001). *Advanced linear modeling: multivariate, time series, and spatial data; nonparametric regression and response surface maximization.* Springer Science & Business Media.

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

Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (5), pp. 851–877.

Cook, R D and Su, Z. (2013). Scaled envelopes: Scale invariant and efficient estimation in multivariate linear regression. *Biometrika* 100 (4), pp. 939–954.

Cook, R. D., Su, Z. and Yang, Y. (2015). envlp: A MATLAB Toolbox for Computing Envelope Estimators in Multivariate Analysis. *Journal of Statistical Software* 62 (1), pp. 1–20.
Cook, R D and Forzani, L and Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika* 102 (2), pp. 439–456.

Cook, R. D., and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association* 110 (510), pp. 599–611.

Cook, R. D., and Zhang, X. (2015). Simultaneous envelopes for multivariate linear regression. *Technometrics* 57 (1), pp. 11–25.

Cook, R. D., and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics* 25 (1), pp. 284–300.

Cook, R. D., Forzani, L., and Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis* 150, pp. 42–54.

Cook, R D and Su, Z. (2016). Scaled Predictor Envelopes and Partial Least-Squares Regression. *Technometrics* 58 (2), pp. 155–165.

Cooper, D. R., Schindler, P. S., and Sun, J. (2003). *Business research methods*. McGraw-Hill/Irwin New York, NY.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

Edelman, A., Arias, T. A., and Smith, S T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications* 20 (2), pp. 303–353.

Guo, Z., Li, L., Lu, W., and Li, B. (2015). Groupwise dimension reduction via envelope method. *Journal of the American Statistical Association* 110 (512), pp. 1515–1527.

Guinness, J., Fuentes, M., Hesterberg, D., and Polizzotto, M. (2014). Multivariate spatial modeling of conditional dependence in microscale soil elemental composition data. *Spatial Statistics* 9, pp. 93–108.

Khare, K., Pal, S. and Su, Z. (2017). A bayesian approach for envelope models. *The Annals of Statistics* 45(1), pp. 196–222.

Kioumourtzoglou, M. A., Schwartz, J. D., Weisskopf, M. G., Melly, S. J., Wang, Y., Dominici, F., and Zanobetti, A. (2016). Long-term PM2.5 exposure and neurological hospital admissions in the northeastern United States. *Environmental Health Perspectives (Online)* 124 (11), pp.23–29.

Lægreid, M and Bockman, O C and Kaarstad, O. (1999). *Agriculture, fertilizers and the environment*. CABI publishing.
Lave, L. B., and Seskin, E. P. (1973). An analysis of the association between US mortality and air pollution. *Journal of the American Statistical Association* 68 (342), pp. 284–290.

Li, L., and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, pp. 1–16.

Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 3–40.

Park, Y and Su, Z and Zhu, H. (2017). Groupwise envelope models for imaging genetic analysis. *Biometrics*. In press.

Phelan, J and Belyazid, S and Jones, P and Cajka, J and Buckley, J and Clark, C. (2016). Assessing the Effects of Climate Change and Air Pollution on Soil Properties and Plant Diversity in Sugar Maple–Beech–Yellow Birch Hardwood Forests in the Northeastern United States: Model Simulations from 1900 to 2100. *Water, Air, & Soil Pollution* 227 (3), pp. 1–30.

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 the American Statistical Association* 81 (393), pp. 142–149.

Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* 98(1), pp. 133–146.

Su, Z. and Cook, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika* 99 (3), pp. 687–702.

Su, Z. and Cook, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica*, pp.213–230.

Su, Z., Zhu, G., Chen, X., and Yang, Y. (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika* 103 (3), pp. 579–593.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media

World Health Organization. (2003). *Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide: report on a WHO working group, Bonn, Germany 13-15 January 2003*. Copenhagen: WHO Regional Office for Europe.
9 Appendix: Theoretical results and prediction plots

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

The likelihood function of the model (8) will be as follows:

\[
L^u(\alpha, \beta^*, V_0, V_1, \theta) = \left[ \det((V_0 + V_1) \otimes \rho(\theta)) \right]^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n - X\beta^*)^T ((V_0 + V_1) \otimes \rho(\theta))^{-1} (Y - \alpha \otimes 1_n - X\beta^*) \right\}
\]

\[
= \left[ \det(V_0 \otimes \rho(\theta) + V_1 \otimes \rho(\theta)) \right]^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n - X\beta^*)^T (V_0 + V_1)^{-1} \otimes \rho^{-1}(\theta) (Y - \alpha \otimes 1_n - X\beta^*) \right\}
\]

\[
= \left[ \det(V_0 \otimes \rho(\theta) + V_1 \otimes \rho(\theta)) \right]^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n - X\beta^*)^T \left( (V_0^\dagger \otimes \rho^{-1}(\theta)) + (V_1^\dagger \otimes \rho^{-1}(\theta)) \right) (Y - \alpha \otimes 1_n - X\beta^*) \right\},
\]

(6)

where \(\dagger\) denotes Moore-Penrose inverse and \(V_0 = \Gamma_0 \Omega_0 \Gamma_0\) and \(V_1 = \Gamma_1 \Omega_1 \Gamma_1\).

Since \(\text{span}(\beta) \subseteq \text{span}(V_1)\) and \(\beta = \Gamma_1 \eta\), therefore we have \(\beta^T = \eta^T \Gamma_1^T\) which means \(\beta^* = \text{vec}(\beta^T) = \text{vec}(\eta^T \Gamma_1^T) = \Gamma_1 \otimes \eta^T \text{vec}(I_u)\).

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

\[
(V_0 \otimes \rho(\theta))\beta^* = (V_0 \otimes \rho(\theta)) (\Gamma_1 \otimes \eta^T) \text{vec}(I_u)
\]

\[
= (V_0 \Gamma_1 \otimes \rho(\theta) \eta^T) \text{vec}(I_u)
\]

\[
= (\Gamma_0 \Omega_0 \Gamma_0 \Gamma_1 \otimes \rho(\theta) \eta^T) \text{vec}(I_u)
\]

\[
= 0,
\]

20
the last equality holds because $\Gamma_1$ and $\Gamma_0$ are orthogonal. Therefore, Since $(V_0 \otimes \rho(\theta))\beta^* = 0$ and $V = V_0 + V_1$, the likelihood in (6) can be factored as:

\[
L(u(\alpha, \beta^*, V_0, V_1, \theta) = \det((V_0 + V_1) \otimes \rho(\theta)) \\
\times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n - X\beta^*)^T \left( (V_1 \otimes \rho^{-1}(\theta)) (Y - \alpha \otimes 1_n - X\beta^*) \right) \right\} \times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n)^T \left( (V_0^1 \otimes \rho^{-1}(\theta)) (Y - \alpha \otimes 1_n) \right) \right\}
\]

\[
= L_1(u(\alpha, \beta^*, V_1, \theta)) \times L_2(u(\alpha, V_0, \theta)),
\]

where $L_1(u(\alpha, \beta^*, V_1, \theta) = \det(V_1) \det(\rho(\theta))^{\frac{\sum}{2}}$}

\[
\times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n - X\beta^*)^T \left( (V_1 \otimes \rho^{-1}(\theta)) (Y - \alpha \otimes 1_n - X\beta^*) \right) \right\},
\]

\[
L_2(u(\alpha, V_0, \theta) = \det(V_0) \det(\rho(\theta))^{\frac{\sum}{2}} \times \exp \left\{ -\frac{1}{2} (Y - \alpha \otimes 1_n)^T \left( (V_0^1 \otimes \rho^{-1}(\theta)) (Y - \alpha \otimes 1_n) \right) \right\},
\]

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

\[
\det((V_0 + V_1) \otimes \rho(\theta)) = \det(V_0 \otimes \rho(\theta) + V_1 \otimes \rho(\theta))
\]

\[
= \det_0(V_0 \otimes \rho(\theta)) + \det_0(V_1 \otimes \rho(\theta))
\]

\[
= [\det_0(V_0)]^{\sum} [\det_0(\rho(\theta))]^{\sum} + [\det_0(V_1)]^{\sum} [\det_0(\rho(\theta))]^{\sum}
\]

\[
= [\det_0(V_0)]^{\sum} [\det(\rho(\theta))]^{\sum} + [\det_0(V_1)]^{\sum} [\det(\rho(\theta))]^{\sum}
\]

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

9.2 Coordinate free version of the algorithm of the spatial envelope

The objective is to maximize the likelihood in (9) over $\alpha, \beta^*, V_0, V_1,$ and $\theta$ subject to the constraints:

\[
\text{span}(\beta) \subseteq \text{span}(V_1), \quad (a)
\]

\[
V_0V_1 = 0, \quad (b).
\]

Based on this factorization given in equation (7), we can decompose the likelihood maximization into the following steps:
1. Fix $\beta, V_0, V_1,$ and $\theta,$ and maximize $L^{(u)}$ in (9) over $\alpha$ which will be:

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

Let $H = Y - \bar{Y} \otimes I_n, U = \text{vec}(H), G = X - \bar{X} \otimes I_n,$ and $F = I_r \otimes G.$ Therefore, the profile likelihood can be written as the following:

$$L_1^u(\beta^*, V_1, \theta) = [\det_0(V_1)]^{-\frac{1}{2}}[\det(\rho(\theta))]^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(U - F\beta^*)^T (V_1^\dagger \otimes \rho^{-1}(\theta))(U - F\beta^*)\right\},$$  \hspace{1cm} (10) 

and

$$L_2^u(V_0, \theta) = [\det_0(V_0)]^{-\frac{1}{2}}[\det(\rho(\theta))]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}U^T (V_0^\dagger \otimes \rho^{-1}(\theta)) U\right\}.$$  \hspace{1cm} (11) 

2. Fix $V_1,$ and $\theta$ and maximize the function $L_1^u$ over $\beta^*,$ subject to (9a), to obtain $L_{21}^u(V_1, \theta).$ Since $\text{vec}(AB) = (I_r \otimes A)\text{vec}(B^T)$ and

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

we have

$$(U - F\beta^*)^T (V_1^\dagger \otimes \rho^{-1}(\theta)) (U - F\beta^*) = tr\left((H - G\beta^T)^T \rho^{-\frac{1}{2}}(\theta)(H - G\beta^T)V_1^\dagger\right)$$

$$= tr\left((H - G\beta^T)^T \rho^{-\frac{1}{2}}(\theta)(H - G\beta^T)\right)$$

$$= tr\left(\rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T\right) V_1^\dagger \left(\rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T\right)^T$$

$$= tr\left(\rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T I_r\right) V_1^\dagger \left(\rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T I_r\right)^T$$

(12)

where $tr(\cdot)$ denotes the trace of the matrix. The last equality in equation (12) is from Lemma 4.1 in Cook et al., (2010). Thus, the optimal $\rho^{-\frac{1}{2}}(\theta)G\beta^T I_r$ is

$$P_{(\rho^{-\frac{1}{2}}(\theta)G)} \left(\rho^{-\frac{1}{2}}(\theta)H\right) P_{(L(V_1))} = P_{(\rho^{-\frac{1}{2}}(\theta)G)} \left(\rho(\theta)^{-\frac{1}{2}}H\right) P_{V_1},$$

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

$$\beta^T = (G^T \rho^{-1}(\theta)G)^{-1} G\rho^{-1}(\theta)HP_{V_1} \Rightarrow \beta = P_{V_1} \hat{\beta},$$

22
where $\beta$ is the MLE estimate of $\beta$ from the full model (8). Substituting this into (11) and using the relation $P_{V_i}V_1 = V_1$, the maximum of $L_2(u)$ for fixed $V_1$ over $\beta$ is

$$L_{11}(V_1, \theta) = [\text{det}(V_1)]^{-\frac{2}{d}} [\text{det}(\rho(\theta))]^{-\frac{2}{d}}$$

$$\times \exp \left\{ -\frac{1}{2} \text{tr} \left( \left( \rho(\theta)^{-\frac{1}{2}}H - P_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)HPV_1} \right) \left( \rho(\theta)^{-\frac{1}{2}}H - P_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)HPV_1} \right)^T \right) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \text{tr} \left( \left( \rho^{-\frac{1}{2}}(\theta)H - P_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} \right) V_1^T \left( \rho^{-\frac{1}{2}}(\theta)H - P_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} \right)^T \right) \right\}$$

$$= [\text{det}(V_1)]^{-\frac{2}{d}} [\text{det}(\rho(\theta))]^{-\frac{2}{d}}$$

where $Q_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} = I_0 - P_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H}$.}

3. Maximize $L^u(V_0, V_1, \theta)$ over all $V_0, V_1$, and $\theta$. Since $L^u(V_0, V_1, \theta) = L_{11}(V_1, \theta) \times L_{22}(V_0, \theta)$, we have

$$L^u(V_0, V_1, \theta) = [\text{det}(V_0)]^{-\frac{2}{d}} [\text{det}(V_1)]^{-\frac{2}{d}} [\text{det}(\rho(\theta))]^{-\frac{2}{d}}$$

$$\times \exp \left\{ -\frac{1}{2} \text{tr} \left( \left( Q_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} V_1 \left( Q_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} \right)^T \right) \right) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \text{tr} \left( \left( Q_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} V_1 \left( Q_{\left(\rho^{-\frac{1}{2}}(\theta)G\right)\rho^{-\frac{1}{2}}(\theta)H} \right)^T \right) \right) \right\}$$

$$= [\text{det}(V_0)]^{-\frac{2}{d}} [\text{det}(V_1)]^{-\frac{2}{d}} [\text{det}(\rho(\theta))]^{-\frac{2}{d}}$$

$$\times \exp \left\{ -\frac{1}{2} \text{tr} \left( \rho^{-\frac{1}{2}}(\theta)HV_0^T \rho^{-\frac{1}{2}}(\theta) \right) \right\}.$$
(b) Fix the $\theta$ and maximize $L^u(V_0, V_1, \theta)$ over $V_0$ and $V_1$. This means maximize $L_{11}^u(V_1, \theta)$ over $V_1$ and $L_{12}^u(V_0, \theta)$ over $V_0$. Maximization $L_{11}^u(P_{V_1})$ over $V_1$ is

$$L_{11}^u(P_{V_1}) \propto \left[ \det_{0} \left( P_{V_1} \left( H^T \rho^{q} \left( Q_{\rho^{-q}(1)G} \right) \rho^{q} \right) H \right) P_{V_1} \right]^{-\frac{\nu}{2}} \tag{15}$$

and maximization $L_{12}^u(P_{V_0})$ over $V_0$ is

$$L_{12}^u(P_{V_0}) \propto \left[ \det_{0} \left( P_{V_0} H^T \rho^{-q}(1)HP_{V_0} \right) \right]^{-\frac{\nu}{2}}. \tag{16}$$

Therefore, maximization $L^u(V_0, V_1, \theta)$ over $V_0$ and $V_1$ is equivalent to maximization of $L_{11}^u(P_{V_1}) \times L_{12}^u(P_{V_0})$ which is proportion to

$$D = \left[ \det_{0} \left( P_{V_1} \left( H^T \rho^{q} \left( Q_{\rho^{-q}(1)G} \right) \rho^{q} \right) H \right) P_{V_1} \right]^{-\frac{\nu}{2}} \times \left[ \det_{0} \left( P_{V_0} H^T \rho^{-q}(1)HP_{V_0} \right) \right]^{-\frac{\nu}{2}}$$

$$= \left[ \det_{0} \left( P_{V_1} \left( H^T \rho^{q} \left( Q_{\rho^{-q}(1)G} \right) \rho^{q} \right) H \right) P_{V_1} + P_{V_0} H^T \rho^{-q}(1)HP_{V_0} \right]^{-\frac{\nu}{2}}$$

$$= \left[ \det_{0} \left( P_{V_1} \left( H^T \rho^{q} \left( Q_{\rho^{-q}(1)G} \right) \rho^{q} \right) H \right) P_{V_1} + Q_{V_0} H^T \rho^{-q}(1)HP_{V_0} \right]^{-\frac{\nu}{2}} \tag{17}$$

where $Q_{V_0} = I_r - P_{V_1}$. Since $\hat{\Sigma}_V = H^T \rho^{-q}(1)H$ and

$$\hat{\Sigma}_{res} = H^T \rho^{q} \left( Q_{\rho^{-q}(1)G} \right) \rho^{q} \right) H$$

$$= H^T \rho^{-q}(1)H$$

$$- H^T \rho^{-q}(1)G \left( G^T \rho^{-q}(1)G \right)^{-1} G^T \rho^{-q}(1)H. \tag{18}$$

Therefore we have $D = \det( P_{V_1} \hat{\Sigma}_{res} P_{V_1} + Q_{V_1} \hat{\Sigma}_V Q_{V_1} )$ and $\hat{V}_1 = \arg\max_{V_1}(D)$ and $P_{V_0} = I_r - P_{V_1}$.

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

### 9.3 Proof of Lemma 1

In this section, we derive the Fisher information matrix for the parameters given by equation (17). Before starting the derivation, the following properties hold:
1. Suppose $A$ and $X$ are both $n \times n$, and $X$ is symmetric, then

$$\frac{\partial \text{vech}(X^{-1})}{\partial \text{vech}(X)^T} = C_n (X^{-1} \otimes X^{-1}) E_n$$

$$\frac{\partial \log(\det(X))}{\partial X} = (\det(X))^{-1} \frac{\partial \det(X)}{\partial X} = \begin{cases} X^{-1} \quad & X \text{ is unconstrained} \\ 2X^{-1} - \text{diag}(X^{-1}) \quad & X \text{ is symmetric} \end{cases}$$

where $C_r \in \mathbb{R}^{r(r+1)/2 \times r^2}$ is expansion matrix which is defined such that for a given matrix such as $A$, $\text{vech}(A) = C_r \text{vec}(A)$, $E_r \in \mathbb{R}^{r \times r(r+1)/2}$ is expansion matrix which is defined such that for a given matrix such as $A$, $\text{vec}(A) = E_r \text{vech}(A)$, and for given square matrix such as $A$, $\text{diag}(A)$ is a matrix with the diagonal elements of $A$.

2. If $Y = AXB$, then

$$\text{tr}(Y) = \text{vec}(A^T B^T) \text{vec}(X) = \text{vec}(A^T B^T) E_n \text{vech}(X),$$

and

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

3. Suppose $B_1$ is an $m \times n$ and $B_2$ is an $n \times q$, matrix, then

$$\text{vec}(B_1 B_2) = (B_2 \otimes I_m) \text{vec}(B_1).$$

4. Suppose $X$ is an $m \times n$ and $A$ is an $n \times n$, matrix, then

$$\frac{\partial \text{vec}(XAX)}{\partial \text{vec}(X)^T} = (X^T A^T \otimes I_n) I_{mn} + (I_n \otimes X^T A).$$

Proof of the above properties can be found in Seber (2008).

The logarithm of the likelihood function (9) is

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

(19)

where $\Theta = \{V, \alpha, \beta^*, \theta\}$. First and second derivatives of the log likelihood function in (19) with respect to $\beta^*$ are
First derivative: \( \frac{\partial \ell(\Theta)}{\partial \beta} = X^T (V^{-1} \otimes \rho^{-1}(\theta))(Y - \alpha \otimes 1_n - XX^T) \)

Second derivative: \( \frac{\partial^2 \ell(\Theta)}{\partial \beta^r \partial \beta^s} = X^T (V^{-1} \otimes \rho^{-1}(\theta))X \)

\[
= (I_n \otimes X^T)(V^{-1} \otimes \rho^{-1}(\theta))(I_n \otimes X)
= V^{-1} \otimes (X^T \rho^{-1}(\theta)X)
\]

From (9), we can rewrite the log likelihood function as

\[
\ell(\Theta) = -\frac{n}{2} \log[\det(V)] - \frac{r}{2} \log[\det(\rho(\theta))]
- \frac{1}{2} tr \left( \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right) V^{-1} \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right)^T \right).
\]

(20)

The \( tr(\cdot) \) is due to

\[
(U - F\beta^*)(V^{-1} \otimes \rho^{-1}(\theta))(U - F\beta^*) = tr \left( (H - G\beta^T)^T \rho^{-1}(\theta)(H - G\beta^T)^T V^{-1} \right)
= tr \left( \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right) V^{-1} \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right)^T \right).
\]

Therefore, the first derivative of the log likelihood function in (20) with respect to \( V \) is

\[
\frac{\partial \ell(\Theta)}{\partial vech(V)} = -\frac{n}{2} vech \left( 2V^{-1} - [\text{diag}(V^{-1})] \right)
- \frac{1}{2} vec \left\{ \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right)^T \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right) \right\} E_r
= -\frac{n}{2} C_r vec \left( 2V^{-1} - [\text{diag}(V^{-1})] \right)
- \frac{1}{2} vec \left\{ \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right)^T \left( \rho^{-\frac{1}{2}}(\theta)H - \rho^{-\frac{1}{2}}(\theta)G\beta^T \right) \right\} E_r
\]

(21)

and second derivative of the log likelihood function in (20) with respect to \( V \) is

\[
\frac{\partial^2 \ell(\Theta)}{\partial vech(V)\partial vech(V)^T} = -n C_r (V^{-1} \otimes V^{-1}) E_r - \frac{n}{2} C_n \left[ \text{diag} \left( V^{-1} \otimes V^{-1} \right) \right] E_r
\]

(22)

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

\[
\frac{\partial^2 \ell(\Theta)}{\partial \text{vech}(\mathbf{V}) \partial \beta^T} = \frac{\partial^2 \ell(\Theta)}{\partial \text{vech}(\mathbf{V}) \partial (\text{vec}(\beta^T))^T}
\]

\[
= -\frac{1}{2} \text{vec}\left\{ \left( \left( \rho^{-\frac{1}{2}}(\theta) \mathbf{H} - \rho^{-\frac{1}{2}}(\theta) \mathbf{G} \beta^T \right)^T \left( \rho^{-\frac{1}{2}}(\theta) \mathbf{H} - \rho^{-\frac{1}{2}}(\theta) \mathbf{G} \beta^T \right) \right) \right\} \mathbf{E}_r
\]

\[
= -\frac{1}{2} \frac{\partial (\text{vec}(\beta^T))^T}{\partial (\text{vec}(\beta^T))^T} \mathbf{E}_r.
\]

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

\[
\text{vec}(\beta \mathbf{G} \rho^{-1}(\theta) \mathbf{H}) = (\mathbf{H}^T \beta \mathbf{G} \otimes \mathbf{I}_r) \text{vec}(\beta)
\]

\[
= (\mathbf{H}^T \beta \mathbf{G} \otimes \mathbf{I}_r) \mathbf{K}_{rp} \text{vec}(\beta^T)
\]

\[
\text{vec}(\mathbf{H}^T \rho^{-1}(\theta) \mathbf{G} \beta^T) = (\mathbf{I}_r \otimes \mathbf{H}^T \rho^{-1}(\theta) \mathbf{G}) \text{vec}(\beta^T).
\]

where \( \mathbf{K}_{rp} \in \mathbb{R}^{p \times p} \) 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 \( \text{vec}(\mathbf{A}^T) = \mathbf{K}_{nn} \text{vec}(\mathbf{A}) \). More properties of \( \mathbf{K}_{nn} \) can be found in Cook, Li, and Chiaromonte (2010) lemma D.2. Therefore, we have

\[
\frac{\text{vec}(\beta \mathbf{G} \rho^{-1}(\theta) \mathbf{H})}{\partial (\text{vec}(\beta^T))^T} = (\mathbf{H}^T \beta \mathbf{G} \otimes \mathbf{I}_r) \mathbf{K}_{rp}
\]

\[
\frac{\text{vec}(\mathbf{H}^T \rho^{-1}(\theta) \mathbf{G} \beta^T) \mathbf{E}_r}{\partial (\text{vec}(\beta^T))^T} = (\mathbf{I}_r \otimes \mathbf{H}^T \rho^{-1}(\theta) \mathbf{G})
\]

\[
\frac{\text{vec}(\beta \mathbf{G}^T \rho^{-1}(\theta) \mathbf{G})}{\partial (\text{vec}(\beta^T))^T} = (\beta \mathbf{G}^T \rho^{-1}(\theta) \mathbf{G} \otimes \mathbf{I}_r) \mathbf{K}_{rp} + (\mathbf{I}_r \otimes \beta \mathbf{G}^T \rho^{-1}(\theta) \mathbf{G}).
\]

Substituting (24) in equation (25), we have

\[
\frac{\partial^2 \ell(\Theta)}{\partial \text{vech}(\mathbf{V}) \partial \beta} = \frac{1}{2} \left\{ (\mathbf{H} - \mathbf{G} \beta^T)^T \rho^{-1}(\theta) \mathbf{G} \otimes \mathbf{I}_r \right\} \mathbf{E}_r
\]

\[
+ \frac{1}{2} \left\{ \mathbf{I}_r \otimes (\mathbf{H} - \mathbf{G} \beta^T)^T \rho^{-1}(\theta) \mathbf{G} \right\} \mathbf{E}_r
\]

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

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

lead to obtain (19).
9.4 Proof of Theorem 1

In this section, we derive the an explicit expression for $\Psi$ as given by (18). In order to find these expression, we need to find expressions for the eight partial derivatives $\frac{\partial \Psi_i}{\partial \phi^T_j}$ for $i = 1, 2$ and $j = 1, 2, 3, 4$.

**Theorem 1:** Suppose $X = 0$ and $J$ is the Fisher information for $\psi(\phi)$ in the model (8):

$$J = \begin{bmatrix} \mathbb{X}^T (V^{-1} \otimes \rho^{-1}(\theta)) \mathbb{X} & 0 \\ 0 & \mathcal{C}_r (V^{-1} \otimes V^{-1}) \mathbf{E}_r - \frac{1}{2} \mathcal{C}_r \left[ \text{diag} (V^{-1} \otimes V^{-1}) \right] \mathbf{E}_r \end{bmatrix}.$$

Then

$$\sqrt{n}(\hat{\phi} - \phi) \rightarrow N(0, \Lambda_0) \quad (26)$$

where $\Lambda_0 = \Psi(\Psi^T \Lambda \Psi)^{-1} \Psi$, $\Lambda = J^{-1}$ is the asymptotic variance of the MLE under the full model, and $\Psi$ is as follows:

$$\begin{bmatrix} \mathbf{K}_{rp}(I_p \otimes \Gamma_1) & \mathbf{K}_{rp}(\eta^T \otimes I_r) & 0 & 0 \\ 0 & 2 \mathcal{C}_r (\Gamma_1 \Omega_1 \otimes I_r - \Gamma_1 \otimes \Gamma_0 \Omega_0 \mathbf{E}_{r-u}) \mathbf{E}_u & \mathcal{C}_r (\Gamma_1 \otimes \Gamma_1) \mathbf{E}_u & \mathcal{C}_r (\Gamma_0 \otimes \Gamma_0) \mathbf{E}_{r-u} \end{bmatrix}.$$

Furthermore, $\Lambda^{-\frac{1}{2}}(\Lambda - \Lambda_0)\Lambda^{-\frac{1}{2}} \succeq 0$, so the spatial envelope model decreases the asymptotic variance.

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

$$\beta^* = \text{vec}(\eta^T \Gamma_1) = \left( \text{vec}(\Gamma_1 \eta) \right)_{(p,r)} = \mathbf{K}_{rp} \text{vec}(\Gamma_1 \eta) = \mathbf{K}_{rp}(I_p \otimes \Gamma_1) \text{vec}(\eta) = \mathbf{K}_{rp}(\eta^T \otimes I_r) \text{vec}(\Gamma_1). \quad (27)$$

Therefore, the derivatives of $\psi_1$ with respect to $\phi^T_1$ is

$$\frac{\partial \psi_1}{\partial \phi^T_1} = \frac{\partial \beta^*}{\partial (\text{vec}(\eta))^T} = \frac{\partial [\mathbf{K}_{rp}(I_p \otimes \Gamma_1) \text{vec}(\eta)]}{\partial (\text{vec}(\eta))^T} = \mathbf{K}_{rp}(I_p \otimes \Gamma_1),$$

and the derivatives of $\psi_1$ with respect to $\phi^T_2$ is

$$\frac{\partial \psi_1}{\partial \phi^T_2} = \frac{\partial \beta^*}{\partial (\text{vec}(\Gamma))^T} = \frac{\partial [\mathbf{K}_{rp}(\eta^T \otimes I_r) \text{vec}(\Gamma_1)]}{\partial (\text{vec}(\Gamma))^T} = \mathbf{K}_{rp}(\eta^T \otimes I_r). \quad (28)$$

It is clear that $\frac{\partial \psi_1}{\partial \phi^T_3} = \frac{\partial \psi_1}{\partial \phi^T_4} = 0$. 

28
The derivative of \( \frac{\partial \psi}{\partial \phi} \) to \( \frac{\partial \psi}{\partial \phi} \) are similar to those in Cook et al. (2010). Having these derivatives together lead to obtain (18).

The asymptotic distribution (26) follows from Shapiro (1986). In order to prove that \( \Lambda_0 \leq \Lambda \), we have

\[
\Lambda_0 - \Lambda = J^{-\frac{1}{2}} - \Psi (\Psi^T \Lambda \Psi)^\dagger \Psi = J^{-\frac{1}{2}} \left[ I_{pr+r(r+1)/2} - J^\frac{1}{2} \Psi (\Psi^T \Lambda \Psi)^\dagger \Psi J^\frac{1}{2} \right] J^{-\frac{1}{2}}
\]

Since the matrix \( I_{pr+r(r+1)/2} - J^\frac{1}{2} \Psi (\Psi^T \Lambda \Psi)^\dagger \Psi J^\frac{1}{2} \) is the projection on to orthogonal complement of \( \text{span}(J^\frac{1}{2} \Psi) \), it is positive semidefinite, which implies that \( \Lambda_0 - \Lambda \) is also positive semidefinite. In addition, we have

\[
\Lambda^{-\frac{1}{2}} (\Lambda - \Lambda_0) \Lambda^{-\frac{1}{2}} = I_{pr+r(r+1)/2} - J^\frac{1}{2} \Psi (\Psi^T \Lambda \Psi)^\dagger \Psi J^\frac{1}{2}
\]

which proves the last statement of the theorem.

9.5 Proof of Corollary 1

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

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

\[
\text{avar}(\sqrt{n}\beta^*) = X^T \left( V^{-1} \otimes \rho(\theta)^{-1} \right) X \otimes \Gamma_1 \Omega_1 \Gamma_1^T + (\eta^T \otimes \Gamma_0) (\Psi_2^T J \Psi_2)^\dagger (\eta \otimes \Gamma_0^T) \tag{29}
\]

where \( \Psi_2 = \left( \frac{\partial \omega_1}{\partial \phi}, \frac{\partial \omega_2}{\partial \phi} \right)^T \).

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

\[
\text{avar}(\sqrt{n}\beta^*) = K_1 (\Psi_1^T J \Psi_1)^\dagger K_1^T + K_1 (\Psi_2^T J \Psi_2)^\dagger K_1^T
\]

where \( \Psi_1 = \left( \frac{\partial \omega_1}{\partial \phi}, \frac{\partial \omega_2}{\partial \phi} \right)^T \), \( K_1 = K_{\text{pr}} (I_p \otimes \Gamma_1) \) and \( K_2 = \eta^T \otimes \Gamma_0 \). Using straightforward matrix multiplication and corollary D1 to D3 in Cook, Li, and Chiaromonte (2010) complete the proof.

9.6 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.

9.7 Prediction Plot for Response Variables
Table 4: Regression Coefficients (asymptotic standard deviation) using envelope the air pollution data in northeastern United States of America.

| 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 5: Regression coefficients (asymptotic standard deviation) using spatial envelope the air pollution data in northeastern United States of America.

| 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 6: Prediction plot of the log of the ground level Ozone for the study area. As it can be seen, the 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 7: Prediction plot of carbon monoxide (CO) for the study area. As it can be seen, the carbon monoxide is moderately low in the study area. CO is high in Rhodes 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 8: Prediction plot of the Nitrogen monoxide for the study area. As it can be seen, the Nitrogen monoxide is high in New York and New Jersey and moderately high almost every place in the study area.

Figure 9: Prediction plot of lead for the study area. As it can be seen, the lead is high in Harrisburg and Lancaster.