Estimation

Model

Suppose the data consist of n observations of the form $\mathcal{D} = \{(\boldsymbol{y}_i, \boldsymbol{x}_{i,1}, \cdots, \boldsymbol{x}_{i,k})\}_{i=1}^n$. Here $\boldsymbol{y}_i \in \mathbb{R}^k$ is a continuous $k \times 1$ response vector, and $\boldsymbol{x}_{i,j}$ is an $p_j \times 1$ vector of covariates for y_{ij}. Conditional on $\mathcal{X}_i = (\boldsymbol{x}_{i,1}, \cdots, \boldsymbol{x}_{i,k})$, the response $\boldsymbol{y}_i$ follows a multivariate normal distribution with unstructured covariance:

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{ik} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{ik} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & & \vdots \\ \vdots & & \ddots & \Sigma_{(k-1)k} \\ \Sigma_{k1} & \cdots & \Sigma_{k(k-1)} & \Sigma_{kk} \end{pmatrix} \right)$$

Regression models for the elements of $\boldsymbol{y}_i$ are given by:

$$\mu_{ij} = E[y_{ij}|\boldsymbol{x}_{i,j}] = \boldsymbol{x}_{i,j}'\boldsymbol{\beta}_j$$

1.1 Notation

Let $\boldsymbol{Y}$ denote the $n \times k$ outcome matrix. The ith row of $\boldsymbol{Y}$ is denoted $\boldsymbol{y}_i$, while the jth column is denoted by $\boldsymbol{t}_j$:

$$\boldsymbol{t}_j \equiv \boldsymbol{Y}[\cdot, j]$$

Let $\boldsymbol{X}_j$ denote the $n \times p_j$ matrix of covariates for $\boldsymbol{t}_j$:

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} \boldsymbol{x}_{1,j}' \\ \vdots \\ \boldsymbol{x}_{n,j}' \end{pmatrix}$$

Let $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ denote the precision matrix.

Likelihood

The log likelihood is:

$$\ell(\boldsymbol{\theta}) \propto -\frac{n}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Lambda}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) \tag{1.2.1}$$

Define $\boldsymbol{V}_i$ as the residual outer product for the ith subject:

$$\boldsymbol{V}_i = (\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'$$

Define the residual matrix:

$$\boldsymbol{E} = \Big(\boldsymbol{t}_1 - \boldsymbol{X}_1\boldsymbol{\beta}_1, \cdots, \boldsymbol{t}_k - \boldsymbol{X}_k\boldsymbol{\beta}_k\Big)$$

Let $\boldsymbol{V}$ denote the summed outer product contributions:

$$\boldsymbol{V} = \sum_{i=1}^{n}\boldsymbol{V}_i = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)' = \boldsymbol{E}'\boldsymbol{E}$$

The log likelihood is expressible as:

$$\ell(\boldsymbol{\theta}) \propto -\frac{n}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Lambda}\boldsymbol{V}\big)$$

Score Equations

3.1 For $\boldsymbol{\beta}_j$

The score equation for $\boldsymbol{\beta}_j$ is:

$$\boldsymbol{U}_j(\boldsymbol{\theta}) = \boldsymbol{X}_j'\boldsymbol{\Lambda}_{jj}(\boldsymbol{t}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j) + \boldsymbol{X}_j'\sum_{l\neq j}\Lambda_{jl}(\boldsymbol{t}_l - \boldsymbol{X}_l\boldsymbol{\beta}_l)$$

3.2 For $\boldsymbol{\Sigma}$

The score equation for $\boldsymbol{\Sigma}$ is:

$$\boldsymbol{U}_{\Sigma} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} - \frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}$$

Estimation Strategy

4.1 Initialization

Initialize $\boldsymbol{\beta}_j$ as:

$$\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{t}_j$$

Construct the initial residual matrix $\boldsymbol{E}^{(0)}$, where the jth column is given by:

$$\boldsymbol{e}_j^{(0)} = \boldsymbol{t}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j^{(0)}$$

Initialize $\boldsymbol{\Sigma}$ using the outer product estimator:

$$\boldsymbol{\Sigma}^{(0)} = n^{-1}\big(\boldsymbol{E}^{(0)}\big)'\boldsymbol{E}^{(0)}$$

4.2 Notation

Let $\boldsymbol{E}_{(-j)}$ denote the residual matrix $\boldsymbol{E}$ with the jth column elided:

$$\boldsymbol{E}_{(-j)} = \boldsymbol{E}[\cdot, -j]$$

Let $\boldsymbol{\Lambda}_{(-j),j}$ denote the $(k-1)\times 1$ column vector obtained by dropping the jth row of $\boldsymbol{\Lambda}$, and subsetting to the jth column:

$$\boldsymbol{\Lambda}_{(-j),j} = \boldsymbol{\Lambda}[-j, j]$$

4.3 Propagation

Procedure 4.1 (**Estimation**)**.** On the rth iteration:

i. Calculate the baseline objective:

$$Q^{(r)} = -\frac{n}{2}\ln\det\boldsymbol{\Sigma}^{(r)}$$

ii. Invert $\boldsymbol{\Sigma}^{(r)}$ to obtain $\boldsymbol{\Lambda}^{(r)}$.

iii. Copy $\boldsymbol{E}^{(r+1)} \leftarrow \boldsymbol{E}^{(r)}$. Note that the residual matrix $\boldsymbol{E}^{(r+1)}$ is updated iteratively as the regression coefficients are updated.

iv. For $j \in \{1, \cdots, k\}$:

(a) Update $\boldsymbol{\beta}_j$ via:

$$\boldsymbol{\beta}_j^{(r+1)} = \boldsymbol{\beta}_j^{(0)} + \big(\boldsymbol{X}_j'\boldsymbol{X}_j\big)^{-1}\boldsymbol{X}_j'\Lambda_{jj}^{-1}\big(\boldsymbol{E}_{(-j)}^{(r+1)}\boldsymbol{\Lambda}_{(-j),j}\big)$$

(b) Update the jth column of $\boldsymbol{E}^{(r+1)}$

$$\boldsymbol{E}_j^{(r+1)} = \boldsymbol{t}_j - \boldsymbol{X}_j\boldsymbol{\beta}^{(r+1)}$$

v. Update $\boldsymbol{\Sigma}^{(r)}$:

$$\boldsymbol{\Sigma}^{(r+1)} = n^{-1}\big(\boldsymbol{E}^{(r+1)}\big)'\boldsymbol{E}^{(r+1)}$$

vi. Calculate the proposed objective:

$$Q^{(r+1)} = -\frac{n}{2} \ln \det \boldsymbol{\Sigma}^{(r+1)}$$

Check the objective for sufficient improvement:

$$\Delta^{(r+1)} = Q^{(r+1)} - Q^{(r)} > \epsilon$$

♠

Inference

Information

1.1 For $\boldsymbol{\beta}_j$

The expected information for $\boldsymbol{\beta}_j$ is:

$$\mathcal{I}_{jj'} = \boldsymbol{X}_j' \Lambda_{jj} \boldsymbol{X}_j$$

The cross information between $\boldsymbol{\beta}_j$ and $\boldsymbol{\beta}_l$ is:

$$\mathcal{I}_{jl'} = \boldsymbol{X}_j' \Lambda_{jl} \boldsymbol{X}_l$$

There is no cross information between the regression and covariance parameters:

$$\mathcal{I}_{j\Sigma_{ab}} = 0$$

Inference on $\boldsymbol{\beta}$

2.1 Partitioning

Fix $\boldsymbol{\beta}_j$ as the regression parameter of interest, and drop the subscript j to reduce notation. Partition $\boldsymbol{\beta} = \big(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B\big)$, and let $\boldsymbol{X} = \big(\boldsymbol{X}_A, \boldsymbol{X}_B\big)$ denote the corresponding partition of the design matrix. Consider the hypothesis $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}$. Group together the nuisance regression parameters as $\boldsymbol{\eta} = (\boldsymbol{\beta}_B, \boldsymbol{\beta}_l)$ where $l \neq j$. Write the joint information of $\boldsymbol{\gamma} = (\boldsymbol{\beta}_A, \boldsymbol{\eta})'$ as:

$$\mathcal{I}_{\gamma\gamma'}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{\beta_A\beta_A'} & \mathcal{I}_{\beta_A\eta'} \\ \mathcal{I}_{\eta\beta_A'} & \mathcal{I}_{\eta\eta'} \end{pmatrix}$$

For $l \neq j$, the component information matrices are:

$$\mathcal{I}_{\beta_A\beta_A'} = \boldsymbol{X}_A' \Lambda_{jj} \boldsymbol{X}_A$$
$$\mathcal{I}_{\beta_A\eta'} = \big(\boldsymbol{X}_A' \Lambda_{jj} \boldsymbol{X}_B, \boldsymbol{X}_A' \Lambda_{jl} \boldsymbol{X}_l\big)$$
$$\mathcal{I}_{\eta\eta'} = \begin{pmatrix} \boldsymbol{X}_B' \Lambda_{jj} \boldsymbol{X}_B & \boldsymbol{X}_B' \Lambda_{jl} \boldsymbol{X}_l \\ \boldsymbol{X}_l' \Lambda_{lj} \boldsymbol{X}_B & \boldsymbol{X}_l' \Lambda_{ll} \boldsymbol{X}_l \end{pmatrix}$$

2.2 Wald Test

The joint distribution of $(\hat{\boldsymbol{\beta}}_A, \hat{\boldsymbol{\eta}})$ is:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger} \\ \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^{\dagger} \end{pmatrix} \dot{\sim} N\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \mathcal{I}_{\beta_A\beta_A'}^{\dagger} & \mathcal{I}_{\beta_A\eta'}^{\dagger} \\ \mathcal{I}_{\eta\beta_A'}^{\dagger} & \mathcal{I}_{\eta\eta'}^{\dagger} \end{pmatrix}^{-1}$$

Define the *efficient information* for $\boldsymbol{\beta}_A$:

$$\mathcal{I}_{\beta_A\beta_A|\eta} \equiv \mathcal{I}_{\beta_A\beta_A'} - \mathcal{I}_{\beta_A\eta'}\mathcal{I}_{\eta\eta'}^{-1}\mathcal{I}_{\eta\beta_A'}$$

Using block inversion, the marginal distribution of $\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger}$ is approximately:

$$\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger} \dot{\sim} N\Big(\mathbf{0},\ \big(\mathcal{I}_{\beta_A\beta_A|\eta}^{\dagger}\big)^{-1}\Big)$$

Here $\mathcal{I}_{\beta_A\beta_A|\eta}^{\dagger}$ denotes evaluation of the efficient information using the true precision $\boldsymbol{\Lambda}^{\dagger}$. The Wald test of $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}$ is:

$$T_W = (\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger})'\mathcal{I}_{\beta_A\beta_A'|\eta}^{\dagger}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger})$$

The realized Wald statistic is:

$$T_W = (\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger})'\hat{\mathcal{I}}_{\beta_A\beta_A'|\eta}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^{\dagger})$$

Here $\hat{\mathcal{I}}_{\beta_A\beta_A|\eta}$ denotes evaluation of the efficient information using the precision $\hat{\boldsymbol{\Lambda}}$ estimated *without* imposing the null hypothesis.

2.3 Score Test

The score equations for $(\boldsymbol{\beta}_A, \boldsymbol{\eta})$ are distributed as:

$$\begin{pmatrix} \boldsymbol{U}_A^{\dagger} \\ \boldsymbol{U}_{\eta}^{\dagger} \end{pmatrix} \dot{\sim} N\begin{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathcal{I}_{\beta_A\beta_A'}^{\dagger} & \mathcal{I}_{\beta_A\eta'}^{\dagger} \\ \mathcal{I}_{\eta\beta_A'}^{\dagger} & \mathcal{I}_{\eta\eta'}^{\dagger} \end{pmatrix} \end{pmatrix}$$

Again $\boldsymbol{U}^{\dagger}$ denotes evaluation of the score using the true regression coefficients $\boldsymbol{\beta}_j^{\dagger}$ and precision $\boldsymbol{\Lambda}^{\dagger}$. The marginal distribution of the score for $\boldsymbol{\beta}_A$ is:

$$\boldsymbol{U}_A^{\dagger} \dot{\sim} N\Big(\mathbf{0},\ \big(\mathcal{I}_{\beta_A\beta_A|\eta}^{\dagger}\big)\Big)$$

The Score test of $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}$ is:

$$T_S = (\boldsymbol{U}_A^{\dagger})'\big(\mathcal{I}_{\beta_A\beta_A'|\eta}^{\dagger}\big)^{-1}\boldsymbol{U}_A^{\dagger}$$

The realized Score statistic is:

$$T_S = \tilde{\boldsymbol{U}}_A'\big(\tilde{\mathcal{I}}_{\beta_A\beta_A'|\eta}\big)^{-1}\tilde{\boldsymbol{U}}_A$$

Here $\tilde{\mathcal{I}}_{\beta_A\beta_A'|\eta}$ denotes evaluation of the efficient information using the precision $\tilde{\boldsymbol{\Lambda}}$ estimated *while* imposing the null hypothesis:

$$\tilde{\boldsymbol{U}}_A = \boldsymbol{U}_A\big(\boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}, \boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}, \boldsymbol{\Lambda} = \tilde{\boldsymbol{\Lambda}}\big)$$

Specifically, $(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\Lambda}})$ satisfy the score equations:

$$\boldsymbol{U}_{\eta}(\boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}, \boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}, \boldsymbol{\Lambda} = \tilde{\boldsymbol{\Lambda}}) = \mathbf{0}$$
$$\boldsymbol{U}_{\Lambda}(\boldsymbol{\beta}_A = \boldsymbol{\beta}_A^{\dagger}, \boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}, \boldsymbol{\Lambda} = \tilde{\boldsymbol{\Lambda}}) = \mathbf{0}$$

Procedure 2.1 (Score Test).

i. Obtain $(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\Lambda}})$ by fitting the model with $\boldsymbol{\beta}_A$ fixed at $\boldsymbol{\beta}_A^{\dagger}$.

ii. Evaluate the score for $\boldsymbol{\beta}_1$ under H_0 as: The score for $\boldsymbol{\beta}_A$ is:

$$\tilde{\boldsymbol{U}}_A = \boldsymbol{X}'_A \tilde{\Lambda}_{jj}(\boldsymbol{t}_j - \boldsymbol{X}_A \boldsymbol{\beta}^{\dagger}_A - \boldsymbol{X}_B \tilde{\boldsymbol{\beta}}_B) + \boldsymbol{X}'_A \sum_{l \neq j} \tilde{\Lambda}_{jl}(\boldsymbol{t}_l - \boldsymbol{X}_l \tilde{\boldsymbol{\beta}}_l)$$

iii. Evaluate the efficient information for $\boldsymbol{\beta}_A$ using $\tilde{\boldsymbol{\Lambda}}$:

$$\tilde{\mathcal{I}}_{\beta_1\beta_1|\eta} = \Big[\mathcal{I}_{\beta_1\beta_1'} - \mathcal{I}_{\beta_1\eta'}\mathcal{I}^{-1}_{\eta\eta'}\mathcal{I}_{\eta\beta_1'}\Big]_{\boldsymbol{\Lambda}=\tilde{\boldsymbol{\Lambda}}}$$

iv. Calculate the score statistic:

$$T_S = \tilde{\boldsymbol{U}}'_A \big(\tilde{\mathcal{I}}_{\beta_A\beta'_A|\eta}\big)^{-1} \tilde{\boldsymbol{U}}_A$$

♠

2.4 Non-centrality

Under the null hypothesis $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}^{\dagger}_A$:

$$\big(\mathcal{I}^{\dagger}_{\beta_A\beta'_A|\eta}\big)^{1/2}\big(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}^{\dagger}_A\big) \xrightarrow{\mathcal{L}} N\big(\boldsymbol{0},\ \boldsymbol{I}\big)$$

Under the sequence of local alternatives $\boldsymbol{\beta}_A = \boldsymbol{\beta}^{\dagger}_A + n^{-1/2}\boldsymbol{\delta}$:

$$\big(\mathcal{I}^{\dagger}_{\beta_A\beta'_A|\eta}\big)^{1/2}\big(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}^{\dagger}_A\big) \xrightarrow{\mathcal{L}} N\Big(\big(\boldsymbol{i}^{\dagger}_{\beta_A\beta'_A|\eta}\big)^{1/2}\boldsymbol{\delta},\ \boldsymbol{I}\Big)$$

Here $\boldsymbol{i}_{\beta_A\beta'_A|\eta}$ is the *unit*, as opposed to the *sample*, efficient information. The non-centrality parameter for the Wald test is therefore:

$$\Delta = \boldsymbol{\delta}'\big(\boldsymbol{i}_{\beta_A\beta'_A|\eta}\big)\boldsymbol{\delta}$$

The estimated non-centrality parameter is exactly the realized Wald statistic $\hat{\boldsymbol{\Delta}} = T_W$.