

## 0.1 Proposed Multivariate Model

Fitting a multivariate LLT allows for the measurement in similarity of underlying constructs for different cognitive tests and compare the effect of some variable across different tests e.g. does APOE have a significantly different effect on cognitive trajectory for the Trails A and Trails B tests? Let  $y_{ijk}$  represent the outcome of cognitive test  $k \in \{1, 2, 3, \dots, K\}$  at time  $j \in \{1, 2, 3, \dots, J\}$  for subject  $i \in \{1, 2, 3, \dots, N\}$ . The model can then be formulated as,

$$\begin{bmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijK} \end{bmatrix} = \begin{bmatrix} \alpha_{ij1} \\ \alpha_{ij2} \\ \vdots \\ \alpha_{ijK} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_{ij}\boldsymbol{\beta}_1 \\ \mathbf{x}_{ij}\boldsymbol{\beta}_2 \\ \vdots \\ \mathbf{x}_{ij}\boldsymbol{\beta}_K \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \vdots \\ \varepsilon_{ijK} \end{bmatrix}, \begin{bmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \vdots \\ \varepsilon_{ijK} \end{bmatrix} \sim N(0, \Sigma_\varepsilon)$$

$$\begin{bmatrix} \alpha_{ij1} \\ \alpha_{ij2} \\ \vdots \\ \alpha_{ijK} \end{bmatrix} = \begin{bmatrix} \alpha_{i(j-1)1} \\ \alpha_{i(j-1)2} \\ \vdots \\ \alpha_{i(j-1)K} \end{bmatrix} + \begin{bmatrix} \eta_{ij1} \\ \eta_{ij2} \\ \vdots \\ \eta_{ijK} \end{bmatrix}, \begin{bmatrix} \eta_{ij1} \\ \eta_{ij2} \\ \vdots \\ \eta_{ijK} \end{bmatrix} \sim N(0, \delta_{ij}\Sigma_\eta)$$

In this model representation,  $\Sigma_\varepsilon \in R^{K \times K}$  and  $\Sigma_\eta \in R^{K \times K}$ . Utilizing  $\Sigma_\varepsilon$  and  $\Sigma_\eta$  allows for estimation of correlation between cognitive tests. However, allowing for non-zero elements in the non-diagonal adds a possibly large number of parameters to estimate. We propose the use of three different models,

Model O: Correlation is in Measurement eq.

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^{2(1)} & \sigma_\varepsilon^{2(1,2)} & \dots & \sigma_\varepsilon^{2(1,K)} \\ \sigma_\varepsilon^{2(1,2)} & \sigma_\varepsilon^{2(2)} & & \\ \vdots & & \ddots & \\ \sigma_\varepsilon^{2(1,K)} & & & \sigma_\varepsilon^{2(K)} \end{bmatrix}, \Sigma_\eta = \begin{bmatrix} \sigma_\eta^{2(1)} & 0 & \dots & 0 \\ 0 & \sigma_\eta^{2(2)} & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_\eta^{2(K)} \end{bmatrix}$$

Model S: Correlation is in State eq.

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^{2(1)} & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^{2(2)} & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_\varepsilon^{2(K)} \end{bmatrix}, \Sigma_\eta = \begin{bmatrix} \sigma_\eta^{2(1)} & \sigma_\eta^{2(1,2)} & \dots & \sigma_\eta^{2(1,K)} \\ \sigma_\eta^{2(1,2)} & \sigma_\eta^{2(2)} & & \\ \vdots & & \ddots & \\ \sigma_\eta^{2(1,K)} & & & \sigma_\eta^{2(K)} \end{bmatrix}$$

Model OS: Correlation is in both Measurement & State eq.

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^{2(1)} & \sigma_\varepsilon^{2(1,2)} & \dots & \sigma_\varepsilon^{2(1,K)} \\ \sigma_\varepsilon^{2(1,2)} & \sigma_\varepsilon^{2(2)} & & \\ \vdots & & \ddots & \\ \sigma_\varepsilon^{2(1,K)} & & & \sigma_\varepsilon^{2(K)} \end{bmatrix}, \Sigma_\eta = \begin{bmatrix} \sigma_\eta^{2(1)} & \sigma_\eta^{2(1,2)} & \dots & \sigma_\eta^{2(1,K)} \\ \sigma_\eta^{2(1,2)} & \sigma_\eta^{2(2)} & & \\ \vdots & & \ddots & \\ \sigma_\eta^{2(1,K)} & & & \sigma_\eta^{2(K)} \end{bmatrix}$$

The correlation allowed between the models makes different assumptions regarding the underlying nature of the test. Model O assumes that any similarity we see in the cognitive scores for a given observation time is due to a correlated random effect on the day of the test. Model S makes the assumption that the underlying cognitive process is correlated between the tests across time. Model OS allows for both types of correlation.

For all models we propose the use of Bayesian estimation to estimate all parameters of interest.

---

**Algorithm 1** Kalman Filter and Smoother for Bayesian Estimation

---

**Kalman Filter**

For  $j$  in  $1, 2, \dots, J$ :

1. Predicted state:  $\alpha_{ij|i(j-1)} = \alpha_{i(j-1)|i(j-1)}$
2. Predicted state variance:  $P_{ij|i(j-1)} = P_{i(j-1)|i(j-1)} + \sigma_\eta^2$
3. Kalman Gain:  $K_{ij} = P_{ij|i(j-1)}(P_{ij|i(j-1)} + \sigma_\epsilon^2)^{-1}$
4. Updated state estimate:  $\alpha_{ij|ij} = \alpha_{ij|i(j-1)} + K_{ij}(\tilde{y}_{ij} - \alpha_{ij|i(j-1)})$
5. Updated state covariance:  $P_{ij|ij} = (1 - K_{ij})P_{ij|i(j-1)}$

**Kalman Smoother**

For  $j^*$  in  $J, J-1, \dots, 1$ :

1. Smoothed predicted state:  $\alpha_{iJ|i(j^*-1)} = \alpha_{i(j^*-1)|i(j^*-1)} + L_{i(j^*-1)}(\alpha_{ij^*|iJ} - \alpha_{ij^*|i(j^*-1)})$
2. Smoothed predicted state variance:  $P_{i(j^*-1)|iJ} = P_{i(j^*-1)|i(j^*-1)} - L_{i(j^*-1)}^2(P_{ij^*|iJ} - P_{ij^*|i(j^*-1)})$

Where  $L_{i(j^*-1)} = P_{i(j^*-1)|i(j^*-1)}P_{ij^*|i(j^*-1)}^{-1}$ .

NOTE: The algorithm is run independently for each  $i \in \{1, 2, \dots, N\}$ .

---

For the underlying process  $\alpha$ , estimation remains similar when compared to [Paper1]. The Kalman Filter independently can be computed for each subject  $i \in \{1, 2, \dots, N\}$ , except all operations are done in matrix form because there is in-subject correlation.

Let  $\psi = \{\Sigma_\eta, \Sigma_\epsilon, \beta\}$ , the vector of the other unknown parameters. Also note, the  $k$  index is dropped and all  $k$ -length parameters are now denoted as vectors. For the posterior of  $\alpha$  we seek to find the distribution defined by,

$$P_\psi(\alpha_{i(0:J)} | \mathbf{y}_{i(1:J)}) = P_\psi(\alpha_{iJ} | \mathbf{y}_{i(1:J)}) P_\psi(\alpha_{i(J-1)} | \alpha_{i(J)}, \mathbf{y}_{1:(J-1)}) \dots P_\psi(\alpha_{i0} | \alpha_{i1})$$

Therefore we need the following densities for  $j$  in  $1, 2, \dots, J-1$  and  $i$  in  $1, 2, \dots, N$ :

$$P_\psi(\alpha_{ij} | \alpha_{i(j+1)}, \mathbf{y}_{i(1:j)}) \propto P_\psi(\alpha_{ij} | \mathbf{y}_{i(1:j)}) P_\psi(\alpha_{i(j+1)} | \alpha_{ij})$$

From the Kalman Filter we calculate  $\alpha_{ij} | \mathbf{y}_{i(1:j)} \sim N_\psi(\alpha_{ij|ij}, P_{ij|ij})$  and  $\alpha_{i(j+1)} | \alpha_{ij} \sim N_\psi(\alpha_{ij}, \Sigma_\eta)$ . After combining the two densities  $\mathbf{m}_{ij} = E_\psi(\alpha_{ij} | \alpha_{i(j+1)}, \mathbf{y}_{i(1:j)}) = \alpha_{ij|ij} + L_{ij}(\alpha_{i(j+1)} - \alpha_{i(j+1)|ij})$  and  $\mathbf{R}_{ij} = \text{Var}_\psi(\alpha_{ij} | \alpha_{i(j+1)}, \mathbf{y}_{i(1:j)}) = P_{ij|ij} - L_{ij}^2 P_{i(j+1)|ij}$  [shumway\_stoffer\_2017]. Because of normality, the posterior distribution for  $\alpha_{ij}$  is  $N(\mathbf{m}_{ij}, \mathbf{R}_{ij})$ .

For the backward sampling procedure we start by sampling  $\alpha_{iJ^*}$  from a  $N_\psi(\mathbf{m}_{iJ}, \mathbf{R}_{iJ})$ , then setting  $\alpha_{iJ^*} = \alpha_{iJ}$  for the calculation of  $\mathbf{m}_{i(J-1)}$  to then sample  $\alpha_{i(J-1)}^*$  from a  $N_\psi(\mathbf{m}_{i(J-1)}, \mathbf{R}_{i(J-1)})$ . This process continues until a whole chain  $\alpha_{i(0:J)}^*$  has been sampled.

Using Bayes rule, the  $\beta$  posterior can be written as follows,

$$P(\beta | Y, \alpha, \Sigma_\eta, \Sigma_\epsilon) = \frac{P(Y, \alpha, \Sigma_\eta, \Sigma_\epsilon | \beta) P(\beta)}{P(Y, \alpha, \Sigma_\eta, \Sigma_\epsilon)}$$

Next, we focus our attention on  $P(Y, \alpha, \Sigma_\eta, \Sigma_\epsilon | \beta)$ ,

$$\begin{aligned}
P(Y, \alpha, \Sigma_\eta, \Sigma_\varepsilon | \beta) &= P(\mathbf{y}_1, \dots, \mathbf{y}_J, \alpha_1, \dots, \alpha_J, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&= P(\mathbf{y}_J | \mathbf{y}_1, \dots, \mathbf{y}_{T-1}, \alpha_1, \dots, \alpha_J, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&\quad \times P(\mathbf{y}_1, \dots, \mathbf{y}_{T-1}, \alpha_1, \dots, \alpha_J, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&= P(\mathbf{y}_J | \alpha_J, \Sigma_\varepsilon | \beta) P(\alpha_{T-1} | \mathbf{y}_1, \dots, \mathbf{y}_{T-1}, \alpha_1, \dots, \alpha_{T-1}, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&\quad \times P(\mathbf{y}_1, \dots, \mathbf{y}_{T-1}, \alpha_1, \dots, \alpha_{T-1}, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&= P(\mathbf{y}_J | \alpha_J, \Sigma_\varepsilon | \beta) P(\alpha_{T-1} | \alpha_{T-1}, \Sigma_\eta) \\
&\quad \times P(\mathbf{y}_1, \dots, \mathbf{y}_{T-1}, \alpha_1, \dots, \alpha_{T-1}, \Sigma_\eta, \Sigma_\varepsilon | \beta) \\
&= P(\Sigma_\varepsilon) P(\Sigma_\eta) \left( \prod_{k=0}^J P(\alpha_k | \alpha_{k-1}, \Sigma_\eta) \right) \prod_{j=1}^J P(\mathbf{y}_j | \alpha_j, \Sigma_\varepsilon, \beta) \\
&\propto \prod_{j=1}^J P(\mathbf{y}_j | \alpha_j, \Sigma_\varepsilon, \beta)
\end{aligned}$$

Due to independence from subject to subject we can write the -2 log likelihood as,

$$-2 \log P(Y, \alpha, \Sigma_\eta, \Sigma_\varepsilon | \beta) \propto \sum_{i=1}^N \sum_{j=1}^T (\mathbf{y}_{ij} - \alpha_{ij} - \mathbb{X}_{ij} \beta)' \Sigma_\varepsilon (\mathbf{y}_{ij} - \alpha_{ij} - \beta)$$

where,

$$\mathbf{y}_{ij} = \begin{bmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijK} \end{bmatrix}, \alpha_{ij} = \begin{bmatrix} \alpha_{ij1} \\ \alpha_{ij2} \\ \vdots \\ \alpha_{ijK} \end{bmatrix}, \mathbb{X}_{ij} = \begin{bmatrix} \mathbf{x}_{ij} & 0 & \dots & 0 \\ 0 & \mathbf{x}_{ij} & & \\ \vdots & & \ddots & \\ 0 & & & \mathbf{x}_{ij} \end{bmatrix} \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

Next, we write the -2 log likelihood in terms of  $\beta$ ,

$$-2 \log P(Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2 | \beta) \propto \beta' \left( \sum_{i=1}^N \sum_{j=1}^T \mathbb{X}_{ij}' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} \right) \beta - 2 \sum_{i=1}^N \sum_{j=1}^T (\mathbf{y}_{ij} - \alpha_{ij})' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} \beta$$

Similarly, the prior can be written in a similar format,

$$\begin{aligned}
-2 \log P(\beta) &\propto (\beta - \theta)' \Sigma_\beta^{-1} (\beta - \theta) \\
&\propto \beta' \Sigma_\beta^{-1} \beta - 2 \theta' \Sigma_\beta^{-1} \beta
\end{aligned}$$

Combining the two we are able to get the proportionality of the posterior,

$$\begin{aligned}
&\beta' \left( \sum_{i=1}^N \sum_{j=1}^T \mathbb{X}_{ij}' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} \right) \beta - 2 \sum_{i=1}^N \sum_{j=1}^T (\mathbf{y}_{ij} - \alpha_{ij})' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} \beta + \beta' \Sigma_\beta^{-1} \beta - 2 \theta' \Sigma_\beta^{-1} \beta \\
&= \beta' \left( \sum_{i=1}^N \sum_{j=1}^T \mathbb{X}_{ij}' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} + \Sigma_\beta^{-1} \right) \beta - 2 \left( \sum_{i=1}^N \sum_{j=1}^T (\mathbf{y}_{ij} - \alpha_{ij})' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} + \theta' \Sigma_\beta^{-1} \right) \beta
\end{aligned}$$

Which, after completing the square, has the form

$$(\beta - \Sigma^{-1} B)' \Sigma (\beta - \Sigma^{-1} B)$$

Where,  $B = \sum_{i=1}^N \sum_{j=1}^T (\mathbf{y}_{ij} - \alpha_{ij})' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} + \theta' \Sigma_\beta^{-1}$  and  $\Sigma = \sum_{i=1}^N \sum_{j=1}^T \mathbb{X}_{ij}' \Sigma_\varepsilon^{-1} \mathbb{X}_{ij} + \Sigma_\beta^{-1}$ .

Thus, the posterior for  $\beta$  has the following distribution,

$$\beta \sim N(\Sigma^{-1}B, \Sigma^{-1})$$

For the posterior of  $\Sigma_\varepsilon$ , if we assume the matrix is non-diagonal as in Model O and OS, then we give  $\Sigma_\varepsilon$  an inverse-Wishart distribution.

$$\begin{aligned} \Sigma_\varepsilon &\sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0) \\ P(\Sigma_\varepsilon) &\propto |\Sigma_\varepsilon|^{-(\nu_0+p+1)/2} \exp(-\text{tr}(\Lambda_0 \Sigma_\varepsilon^{-1})) \\ P(Y, \alpha, \Sigma_\eta, \beta | \Sigma_\varepsilon) &\propto |\Sigma_\varepsilon|^{-(NJ)/2} \exp(-\sum_{i=1}^N \sum_{j=1}^J (\mathbf{y}_{ij} - \alpha_{ij} - \mathbb{X}_{ij}\beta)' \Sigma_\varepsilon^{-1} (\mathbf{y}_{ij} - \alpha_{ij} - \mathbb{X}_{ij}\beta) / 2) \\ P(\Sigma_\varepsilon | Y, \alpha, \Sigma_\eta, \beta) &= P(Y, \alpha, \Sigma_\eta, \beta | \Sigma_\varepsilon) P(\Sigma_\varepsilon) \\ &\propto |\Sigma_\varepsilon|^{-(NJ+\nu_0+p+1)/2} \exp(-\text{tr}(\Lambda_0 + \sum_{i=1}^N \sum_{j=1}^J (\mathbf{y}_{ij} - \alpha_{ij} - \mathbb{X}_{ij}\beta)^2 \Sigma_\varepsilon^{-1})) \\ \Sigma_\varepsilon &\sim \text{Inv-Wishart}_{\nu_0+NJ}(\Lambda_0 + \sum_{i=1}^N \sum_{j=1}^J (\mathbf{y}_{ij} - \alpha_{ij} - \mathbb{X}_{ij}\beta)^2) \end{aligned}$$

The posterior for an assumed diagonal matrix  $\Sigma_\varepsilon$  can be calculated independently for each test as was previously shown [Paper1] using the inverse-gamma for each test.

On the other hand, if  $\Sigma_\eta$  is non-diagonal as was shown in Model O and OS,

$$\begin{aligned} \Sigma_\eta &\sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0) \\ P(\Sigma_\eta) &\propto |\Sigma_\eta|^{-(\nu_0+p+1)/2} \exp(-\text{tr}(\Lambda_0 \Sigma_\eta^{-1})) \\ P(Y, \alpha, \Sigma_\eta, \beta | \Sigma_\eta) &\propto |\Sigma_\eta|^{-(N(J-1))/2} \exp(-\sum_{i=1}^N \sum_{j=2}^J (\alpha_{ij} - \alpha_{i(j-1)})' \Sigma_\eta^{-1} (\alpha_{ij} - \alpha_{i(j-1)}) / 2) \\ P(\Sigma_\eta | Y, \alpha, \Sigma_\eta, \beta) &= P(Y, \alpha, \Sigma_\eta, \beta | \Sigma_\eta) P(\Sigma_\eta) \\ &\propto |\Sigma_\eta|^{-(N(J-1)+\nu_0+p+1)/2} \exp(-\text{tr}(\Lambda_0 + \sum_{i=1}^N \sum_{j=2}^J (\alpha_{ij} - \alpha_{i(j-1)})^2 \Sigma_\eta^{-1})) \\ \Sigma_\eta &\sim \text{Inv-Wishart}_{\nu_0+N(J-1)}(\Lambda_0 + \sum_{i=1}^N \sum_{j=2}^J (\alpha_{ij} - \alpha_{i(j-1)})^2) \end{aligned}$$

Once again, the posterior for an assumed diagonal matrix  $\Sigma_\eta$  can be calculated independently for each test as was previously shown [Paper1].

The covariance matrix  $\Sigma_\varepsilon$  and  $\Sigma_\eta$ , both offer meaningful insight into underlying constructs of cognition. Additionally, these models allow for comparison of linear effects across different tests.

In order to compare effects across cognition test, the test must be standardized. However, ad-hoc outcome standardization and joint-modeling using the SSM models does not allow for direct comparison of effects across tests. This is because it would not take into account what amount of variance is coming from the observation equation and what amount of variance is coming from the underlying temporal cognition process. To standardize the outcome we instead rely on estimating  $\Sigma_\varepsilon^{(m)}$  at each iteration  $m \in \{1, 2, \dots, M\}$ , then multiplying  $\tilde{\beta}^{(m)} = [\beta_1^{(m)} \dots \beta_K^{(m)}]$  with  $(\Sigma_\varepsilon^{(m)})^{-1}$ . This will yield a standardized effect estimate for each test that can then be compared to on another.

At iteration  $m$  our model is as follows,

$$\begin{aligned}
\mathbf{y}_{ij} &= \boldsymbol{\alpha}_{ij}^{(m)} + \mathbb{X}_{ij}\boldsymbol{\beta}^{(m)} + \boldsymbol{\varepsilon}_{ij} \\
\mathbf{y}_{ij} - \boldsymbol{\alpha}_{ij}^{(m)} &= \mathbb{X}_{ij}\boldsymbol{\beta}^{(m)} + \boldsymbol{\varepsilon}_{ij} \\
(\mathbf{y}_{ij} - \boldsymbol{\alpha}_{ij}^{(m)})(\Sigma_{\varepsilon}^{(m)})^{-1} &= \mathbb{X}_{ij}\boldsymbol{\beta}^{(m)}(\Sigma_{\varepsilon}^{(m)})^{-1} + \boldsymbol{\varepsilon}_{ij}(\Sigma_{\varepsilon}^{(m)})^{-1} \\
\mathbf{y}_{ij}^{(m)*} &= \mathbb{X}_{ij}\boldsymbol{\beta}^{(m)}(\Sigma_{\varepsilon}^{(m)})^{-1} + \boldsymbol{\varepsilon}_{ij}^*
\end{aligned}$$

Where  $\mathbf{y}_{ij}^{(m)*} \sim N(0, I)$ , meaning each test has the same observed variance 1 and each tend has had the trend component removed. We then let  $\boldsymbol{\beta}^{(m)*} = \boldsymbol{\beta}^{(m)}(\Sigma_{\varepsilon}^{(m)})^{-1}$  which is a standardized linear effect estimate and the respective components can be used for hypothesis testing across tests.