Derivation of system of equations

Emily C. Voldal, Fan Xia, Avi Kenny, Patrick J. Heagerty, and James P. Hughes

Contents

1 Introduction and notes

This document provides mathematical details on the derivation of the system of equations from the paper 'Model misspecification in stepped wedge trials: Random effects for time or treatment'. To improve readability of long equations, some simplifications have been made to notation presented in the paper. For example, subscripts indicating sequence-specific values are mostly suppressed. To smooth 'translation' between notations, we will re-present the models and cases of interest. Throughout, $diag(a, \dots, b)$ represents a diagonal matrix with $a, \dots, b$ on the diagonal and zeros elsewhere. $\mathbf{1}_a$ and $\mathbf{0}_a$ represent matrices with a rows and one column, filled with ones and zeros, respectively. I_a represents an identity matrix of dimension a.

2 The General Model

We are considering a stepped wedge trial that has a total of J time periods and K individual observations per cluster per time period (cross-sectional design). We assume that every cluster is observed at each time period, and that once a cluster crosses over to treatment it remains on treatment for the duration of the trial. We also assume there are an equal number of clusters in each sequence. These assumptions allow for a wide variety of classical and non-classical designs, as long as each sequence consists of some number of time points on control followed by some number of time points on treatment; one notable design element which breaks this assumption is transition periods, where outcomes are not observed for one or more time points after crossover. See supplemental R code for more examples of acceptable non-classical designs. The results in this document (i.e. the system of equations and its roots) are valid for any reasonable set of fixed effects, including different methods of modeling time (e.g. linear categorical, splines).

For a specific sequence, T (suppressing index) is the total number of time periods on treatment. For example, if a sequence crossed over after the second time point in a study with five time points, $T = 3$ since it spent the first two periods on control and the last three periods on treatment.

Generally, we will be considering a mixed model of the following form, where outcomes for cluster i are represented by:

$$Y_i = X_i\theta + Z_i a_i + \epsilon_i$$

where $X_i\theta$ represents the fixed effects, $Z_i a_i$ represents the random effects, and ϵ_i represents the residual error. The vector of coefficients for the fixed effects θ should include at least an intercept, treatment effect, and some way to model time. For example, $[\mu, \beta, \theta]^T$ was used in the main paper (note that θ represented the treatment effect in the main paper, but now we are using it to represent the whole vector of fixed effects). The corresponding design matrix X_i has JK rows and an appropriate number of columns. Additionally,

$$
\begin{aligned}
Y_i &= [Y_{i11}, Y_{i12}, \ldots, Y_{iJK}]^T, dim(JK,1) \\
\epsilon_i &= [\epsilon_{i11}, \ldots, \epsilon_{iJK}]^T, dim(JK,1)
\end{aligned}
$$

3 Time-fitted random treatment case

In this case, we are fitting a model with random time and intercept effects, but in truth the data comes from a model with random treatment and intercept effects.

3.1 Misspecified model

The misspecified model with random time effects is presented below. The fixed effects remain correctly specified (see Section 2).

$$
Y_i = X_i\theta + Z_i a_i + \epsilon_i
$$

$$
\begin{aligned}
Z_i &= [\mathbf{1}_{JK}, \begin{matrix}\mathbf{1}_K \\ \mathbf{0}_{JK-K}\end{matrix}, \ldots, \begin{matrix}\mathbf{0}_{JK-K} \\ \mathbf{1}_K\end{matrix}], dim(JK, J+1) \\
a_i &= [u_i, w_{i1}, \ldots, w_{iJ}]^T \sim MVN(0, G), dim(J+1, 1) \\
G &= diag(\tau^2, \gamma^2, \ldots, \gamma^2), dim(J+1, J+1) \\
\epsilon_i &\sim MVN(0, \sigma^2 I_{JK}), dim(JK, 1)
\end{aligned}
$$

3.2 True model

The correctly specified model with random treatment effects is presented below. Note the 't' subscript, indicating that the matrix or parameter comes from the true model.

$$
Y_i = X_i\theta + Z_{it} a_{it} + \epsilon_{it}
$$

$$
\begin{aligned}
Z_{it} &= [\mathbf{1}_{JK}, \begin{matrix}\mathbf{0}_{K(J-T_i)} \\ \mathbf{1}_{KT_i}\end{matrix}], dim(JK, 2) \\
a_{it} &= [u_{it}, v_{it}]^T \sim MVN(0, G_t), dim(2, 1) \\
G_t &= diag(\tau_t^2, \eta_t^2), dim(2, 2) \\
\epsilon_{it} &\sim MVN(0, \sigma_t^2 I_{JK}), dim(JK, 1)
\end{aligned}
$$

3.3 Marginal misspecified likelihood

Recall that the equation we are trying to find roots for (Equation 3 in the main paper) is a sum over sequences, since the marginal likelihood is the same for every cluster within a sequence. For the next several sections, we'll be doing computations for a single sequence - that is, a single cluster within a sequence. To improve readability, the i subscripts in the notation presented above will be repressed.

Since $Y|a, X, Z \sim MVN(X\theta + Za, \sigma^2 I)$ under the mis-specified model, the conditional mis-specified likelihood is:

$$\begin{aligned} Pr(Y|a, X, Z) &= (2\pi)^{-JK/2}|\sigma^2 I|^{-1/2} exp\{-\frac{1}{2}(Y - (X\theta + Za))^T(\sigma^2 I)^{-1}(Y - (X\theta + Za))\} \\ &= (2\pi\sigma^2)^{-JK/2} exp\{-\frac{1}{2\sigma^2}(Y - (X\theta + Za))^T(Y - (X\theta + Za))\} \end{aligned}$$

Next, we integrate over the random effects to get the marginal mis-specified likelihood.

$$\begin{aligned} Pr(Y|X, Z) &= \int Pr(Y|a, X, Z)f(a|X, Z)da \qquad (1) \\ &= \int (2\pi\sigma^2)^{-JK/2} exp\{-\frac{1}{2\sigma^2}((Y - X\theta) - Za)^T((Y - X\theta) - Za)\} \\ &\quad (2\pi)^{-(J+1)/2}|G|^{-1/2} exp\{-\frac{1}{2}a^T G^{-1} a\}da \\ &= (2\pi\sigma^2)^{-JK/2}(2\pi)^{-(J+1)/2}(\tau^2\gamma^{2J})^{-1/2} exp\{-\frac{1}{2\sigma^2}(Y - X\theta)^T(Y - X\theta)\} \\ &\quad \int exp\{-\frac{1}{2}(\frac{1}{\sigma^2}(-a^T Z^T(Y - X\theta) - (Y - X\theta)^T Za) + a^T(Z^T Z\frac{1}{\sigma^2} + G^{-1})a)\}da \end{aligned}$$

We wish to solve the integral by completing the square and obtaining a MVN($\frac{1}{\sigma^2}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta)$,$(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}$) pdf for a. Below are some useful facts about this covariance matrix.

$$\begin{aligned} Z^T Z\frac{1}{\sigma^2} + G^{-1} &= \frac{1}{\sigma^2}\begin{bmatrix} JK & K\mathbf{1}_J^T \\ K\mathbf{1}_J & KI_J \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau^2} & \mathbf{0}_J^T \\ \mathbf{0}_J & \frac{1}{\gamma^2}I_J \end{bmatrix} \\ &= \begin{bmatrix} \frac{JK}{\sigma^2} + \frac{1}{\tau^2} & \frac{K}{\sigma^2}\mathbf{1}_J^T \\ \frac{K}{\sigma^2}\mathbf{1}_J & (\frac{K}{\sigma^2} + \frac{1}{\gamma^2})I_J \end{bmatrix} \end{aligned}$$

Since the determinant of the inverse is the inverse of the determinant, we can skip inverting for now (note that this matrix is invertible and symmetric). Since the matrix has a 2x2 block structure, we can find the determinant in the following way:

$$|Z^TZ\frac{1}{\sigma^2}+G^{-1}| = |(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})I_J|(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}-\frac{K}{\sigma^2}\mathbf{1}_J^T((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})I_J)^{-1}\frac{K}{\sigma^2}\mathbf{1}_J)$$
$$= (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^J(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{J-1}$$

So, multiplying and dividing Equation 1 by the term below

$$|(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}|^{-1/2}exp\{\frac{-1}{2}(\frac{1}{\sigma^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta))^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})$$
$$(\frac{1}{\sigma^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta))\}$$

we get:

$$Pr(Y|X,Z) = (2\pi\sigma^2)^{-JK/2}(\tau^2\gamma^{2J})^{-1/2}exp\{-\frac{1}{2\sigma^2}(Y-X\theta)^T(Y-X\theta)\}$$
$$((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^J(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{J-1})^{-1/2}$$
$$exp\{\frac{1}{2}(\frac{1}{\sigma^2})^2((Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta))^T(Z^TY-Z^TX\theta)\}$$
$$= (2\pi\sigma^2)^{-JK/2}(\tau^2\gamma^{2J})^{-1/2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^J(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{J-1})^{-1/2}$$
$$exp\{-\frac{1}{2\sigma^2}(Y-X\theta)^T(Y-X\theta)$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY-Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta)\}$$

So the log marginal likelihood is:

$$log(Pr(Y|X,Z)) = -\frac{JK}{2}log(2\pi\sigma^2) - \frac{1}{2}log(\tau^2\gamma^{2J}) \tag{2}$$
$$-\frac{1}{2}log((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)$$
$$-\frac{J-1}{2}log(\frac{K}{\sigma^2}+\frac{1}{\gamma^2}) - \frac{1}{2\sigma^2}(Y-X\theta)^T(Y-X\theta)$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY-Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta)$$

3.4 Score equations

Fixed effects

Note that we are still keeping all the fixed effects in the vector θ, so this score equation is a vector also.

$$\frac{\partial}{\partial\theta}log(Pr(Y|X,Z)) = -\frac{1}{\sigma^2}(-X^TY + X^TX\theta) - (\frac{1}{\sigma^2})^2X^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)$$

Residual variance

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z)) = &-\frac{JK}{2}\frac{1}{\sigma^2} + \frac{1}{2}((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})(\frac{JK}{\sigma^2} + \frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)^{-1}(\frac{K}{(\sigma^2)^2})(\frac{J}{\gamma^2} + \frac{1}{\tau^2})\\ &+ \frac{J-1}{2}(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1}\frac{K}{(\sigma^2)^2} + \frac{1}{2(\sigma^2)^2}(Y - X\theta)^T(Y - X\theta)\\ &- (\frac{1}{\sigma^2})^3(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)\\ &+ \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T[\frac{\partial}{\sigma^2}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)\end{aligned}$$

Note that $\frac{\partial}{\partial x}M(x)^{-1} = -M(x)^{-1}[\frac{\partial}{\partial x}M(x)]M(x)^{-1}$ and $\frac{\partial}{\sigma^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = Z^TZ\frac{-1}{(\sigma^2)^2}$, so

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z)) = &-\frac{JK}{2}\frac{1}{\sigma^2} + \frac{1}{2}((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})(\frac{JK}{\sigma^2} + \frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)^{-1}(\frac{K}{(\sigma^2)^2})(\frac{J}{\gamma^2} + \frac{1}{\tau^2})\\ &+ \frac{J-1}{2}(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1}\frac{K}{(\sigma^2)^2} + \frac{1}{2(\sigma^2)^2}(Y - X\theta)^T(Y - X\theta)\\ &- (\frac{1}{\sigma^2})^3(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)\\ &+ \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ\\ &\frac{1}{(\sigma^2)^2}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)\end{aligned}$$

Random effect variances

Similarly, because $\frac{\partial}{\tau^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = diag\{\frac{-1}{(\tau^2)^2}, 0, \ldots, 0\}$ and $\frac{\partial}{\gamma^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = diag\{0, \frac{-1}{(\gamma^2)^2}, \ldots, \frac{-1}{(\gamma^2)^2}\}$,

$$\frac{\partial}{\partial \tau^2} log(Pr(Y|X,Z)) = -\frac{1}{2}\frac{1}{\tau^2} + \frac{1}{2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)^{-1}(\frac{1}{(\tau^2)^2})(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1} diag\{\frac{1}{(\tau^2)^2}, 0, \ldots, 0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY - Z^TX\theta)$$
$$\frac{\partial}{\partial \gamma^2} log(Pr(Y|X,Z)) = -\frac{J}{2}\frac{1}{\gamma^2} + \frac{1}{2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)^{-1}(\frac{1}{(\gamma^2)^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2})$$
$$+\frac{J-1}{2}(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{-1}(\frac{1}{(\gamma^2)^2}) + \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1} diag\{0, \frac{1}{(\gamma^2)^2}, \ldots, \frac{1}{(\gamma^2)^2}\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY - Z^TX\theta)$$

3.5 Expectations

Next, we take expectations of the score equations, with respect to the true distribution of Y (that is, the correctly specified model). In this case (Normal outcomes, identity link), we know that θ is unbiased, so $E(Y|X, Z_t) = E(E(Y|X, Z_t, a_t)) = E(X\theta + Z_t a_t) = X\theta$. We will also make frequent use of the rule $E(x^TAx) = tr(Acov(x)) + E(x)^TAE(x)$.

Because it occurs so frequently, we will use $\xi = (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2$ as shorthand.

Many of these expectations involve finding the trace of products of matrices. All the matrices involved have a convenient block structure, either 2-by-2 blocks or 3-by-3 blocks. Most of the matrix multiplication work is straightforward but lengthy, so is not described explicitly. However, to aid understanding some useful facts which are used many times in the calculations are presented below.

$$(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1} = \begin{bmatrix} \frac{JK}{\sigma^2}+\frac{1}{\tau^2} & \frac{K}{\sigma^2}\mathbf{1}_J^T \\ \frac{K}{\sigma^2}\mathbf{1}_J & (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})I_J \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})\frac{1}{\xi} & -\frac{K}{\sigma^2\xi}\mathbf{1}_J^T \\ -\frac{K}{\sigma^2\xi}\mathbf{1}_J & (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{-1}I_J + (\frac{K}{\sigma^2})^2((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})\xi)^{-1}\mathbf{1}_J\mathbf{1}_J^T \end{bmatrix}$$

$$(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^TZ = \begin{bmatrix} (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})\frac{1}{\xi} & -\frac{K}{\sigma^2\xi}\mathbf{1}_J^T \\ -\frac{K}{\sigma^2\xi}\mathbf{1}_J & (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{-1}I_J + (\frac{K}{\sigma^2})^2((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})\xi)^{-1}\mathbf{1}_J\mathbf{1}_J^T \end{bmatrix}\begin{bmatrix} JK & K\mathbf{1}_J^T \\ K\mathbf{1}_J & KI_J \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\gamma^2}\frac{1}{\xi}JK & \frac{1}{\gamma^2}\frac{1}{\xi}K\mathbf{1}_J^T \\ \frac{1}{\tau^2}\frac{1}{\xi}K\mathbf{1}_J & (\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{-1}KI_J - K\frac{K}{\sigma^2}\frac{1}{\gamma^2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})\xi)^{-1}\mathbf{1}_J\mathbf{1}_J^T \end{bmatrix}$$

$$Z^T Z_t G_t Z_t^T Z = \begin{bmatrix} JK & TK \\ K\mathbf{1}_{J-T} & \mathbf{0}_{J-T} \\ K\mathbf{1}_T & K\mathbf{1}_T \end{bmatrix} \begin{bmatrix} \tau_t^2 & 0 \\ 0 & \eta_t^2 \end{bmatrix} \begin{bmatrix} JK & TK \\ K\mathbf{1}_{J-T} & \mathbf{0}_{J-T} \\ K\mathbf{1}_T & K\mathbf{1}_T \end{bmatrix}^T$$
$$= \begin{bmatrix} J^2K^2\tau_t^2 + T^2K^2\eta_t^2 & JK^2\tau_t^2\mathbf{1}_{J-T}^T & (JK^2\tau_t^2 + TK^2\eta_t^2)\mathbf{1}_T^T \\ JK^2\tau_t^2\mathbf{1}_{J-T} & K^2\tau_t^2\mathbf{1}_{J-T}\mathbf{1}_{J-T}^T & K^2\tau_t^2\mathbf{1}_{J-T}\mathbf{1}_T^T \\ (JK^2\tau_t^2 + TK^2\eta_t^2)\mathbf{1}_T & K^2\tau_t^2\mathbf{1}_T\mathbf{1}_{J-T}^T & K^2(\tau_t^2 + \eta_t^2)\mathbf{1}_T\mathbf{1}_T^T \end{bmatrix}$$

Fixed effects

$$\begin{aligned} E[\frac{\partial}{\partial\theta} log(Pr(Y|X,Z)] &= -\frac{1}{\sigma^2}(-X^T E[Y] + X^T X\theta) - (\frac{1}{\sigma^2})^2 X^T Z(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T E[Y] - Z^T X\theta) \\ &= -\frac{1}{\sigma^2}(-X^T X\theta + X^T X\theta) - (\frac{1}{\sigma^2})^2 X^T Z(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T X\theta - Z^T X\theta) \\ &= \mathbf{0} \end{aligned}$$

Residual variance

$$\begin{aligned} E[\frac{\partial}{\partial\sigma^2} log(Pr(Y|X,Z))] = &-\frac{JK}{2}\frac{1}{\sigma^2} \\ &+ \frac{1}{2}((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})(\frac{JK}{\sigma^2} + \frac{1}{\tau^2}) - J(\frac{K}{\sigma^2})^2)^{-1}(\frac{K}{(\sigma^2)^2})(\frac{J}{\gamma^2} + \frac{1}{\tau^2}) \\ &+ \frac{J-1}{2}(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1}\frac{K}{(\sigma^2)^2} + \frac{1}{2(\sigma^2)^2}E[(Y - X\theta)^T(Y - X\theta)] \\ &- (\frac{1}{\sigma^2})^3 E[(Z^T Y - Z^T X\theta)^T(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta)] \\ &+ \frac{1}{2}(\frac{1}{\sigma^2})^4 E[(Z^T Y - Z^T X\theta)^T \\ &[(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1} Z^T Z(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^T Y - Z^T X\theta)] \end{aligned} \tag{3}$$

To solve the first expectation in Equation 3, note that

$$\begin{aligned} E[(Y - X\theta)^T(Y - X\theta)] = tr\{cov(Y)\} &= tr\{Z_t G_t Z_t^T + \sigma_t^2 I_{JK}\} \\ &= (J-T)K\tau_t^2 + TK(\tau_t^2 + \eta_t^2) + JK\sigma_t^2 \\ &= JK\tau_t^2 + TK\eta_t^2 + JK\sigma_t^2 \end{aligned}$$

To solve the second expectation in Equation 3, note that

$$E[(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)] = tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}cov(Z^TY)\}$$
$$= tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^T(Z_tG_tZ_t^T + \sigma_t^2 I_{JK})Z\}$$
$$= tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ_tG_tZ_t^TZ\}$$
$$+ \sigma_t^2 tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ\}$$
$$= \frac{1}{\gamma^2}\frac{1}{\xi}(J^2K^2\tau_t^2 + T^2K^2\eta_t^2) + (J-T)K^2\tau_t^2\frac{1}{\tau^2\xi}$$
$$+ T\frac{-K}{\sigma^2}\frac{1}{\gamma^2}((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})\xi)^{-1}(JK^2\tau_t^2 + TK^2\eta_t^2) + T(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1}K^2(\tau_t^2 + \eta_t^2)$$
$$+ \sigma_t^2[JK\sigma^2(\gamma^2\xi - \frac{K}{\sigma^2})(\xi(\gamma^2K + \sigma^2))^{-1} + \frac{JK}{\gamma^2\xi}]$$

To solve the third expectation in Equation 3, note that

$$E[(Z^TY - Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]$$
$$= tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ_tG_tZ_t^TZ\}$$
$$+ \sigma_t^2 tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ\}$$
$$= [J\tau_t^2 + T\eta_t^2][K^3(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-2}]$$
$$+ [J^2K^2\tau_t^2 + T^2K^2\eta_t^2][\frac{1}{\tau^2}\frac{1}{\xi^2}K(\frac{1}{\gamma^2})^2(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1} - \frac{K}{\sigma^2}\frac{1}{\xi}(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-2}(\frac{1}{\gamma^2})K$$
$$+ (\frac{1}{\gamma^2}\frac{1}{\xi})^2JK + \frac{K}{\gamma^2\tau^2\xi^2}] + \sigma_t^2[(\frac{1}{\gamma^2}\frac{1}{\xi})^2J^2K^2 + \frac{JK^2}{\gamma^2\tau^2\xi^2} + J((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-2}K^2$$
$$+ \frac{1}{\tau^2}\frac{1}{\xi^2}K^2(\frac{1}{\gamma^2}) - \frac{K}{\sigma^2}\frac{1}{\xi}(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-1}K^2 + JK^2(\frac{K}{\sigma^2})^2\frac{1}{\gamma^2}((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})\xi^2)^{-1}$$
$$+ K^2(\frac{K}{\sigma^2})^2((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^2\xi)^{-1} - K^2\frac{K}{\sigma^2}\frac{1}{\gamma^2}(\frac{JK}{\sigma^2} + \frac{1}{\tau^2})((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})\xi^2)^{-1})]$$

Random effect variances
We begin with the score equation for the random intercept effect.

$$
\begin{aligned}
E[\frac{\partial}{\partial\tau^2}log(Pr(Y|X,Z))] = &-\frac{1}{2}\frac{1}{\tau^2}+\frac{1}{2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2})-J(\frac{K}{\sigma^2})^2)^{-1}(\frac{1}{(\tau^2)^2})(\frac{K}{\sigma^2}+\frac{1}{\gamma^2}) \\
&+\frac{1}{2}(\frac{1}{\sigma^2})^2(\frac{1}{(\tau^2)^2})E[(Z^TY-Z^TX\theta)^T \\
&[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{1,0,\ldots,0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)]
\end{aligned}
$$

Note that the one expectation in this equation is:

$$
\begin{aligned}
&E[(Z^TY-Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{1,0,\ldots,0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)] \\
&\quad = tr\{[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{1,0,\ldots,0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}]Z^TZ_tG_tZ_t^TZ\} \\
&\quad\quad + \sigma_t^2 tr\{[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{1,0,\ldots,0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}]Z^TZ\} \\
&\quad = (\frac{1}{\gamma^2})^2\frac{1}{\xi^2}(J^2K^2\tau_t^2+T^2K^2\eta_t^2)+\sigma_t^2JK\frac{1}{\xi^2}(\frac{1}{\gamma^2})^2
\end{aligned}
$$

The score equation for the random time effect has a similar form.

$$
\begin{aligned}
E[\frac{\partial}{\partial\gamma^2}log(Pr(Y|X,Z))] = &-\frac{J}{2}\frac{1}{\gamma^2}+\frac{1}{2}((\frac{K}{\sigma^2}+\frac{1}{\gamma^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2})-J(\frac{K}{\sigma^2})^2)^{-1}(\frac{1}{(\gamma^2)^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) \\
&+\frac{J-1}{2}(\frac{K}{\sigma^2}+\frac{1}{\gamma^2})^{-1}(\frac{1}{(\gamma^2)^2})+\frac{1}{2}(\frac{1}{\sigma^2})^2\frac{1}{(\gamma^2)^2}E[(Z^TY-Z^TX\theta)^T \\
&[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{0,1,\ldots,1\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)]
\end{aligned}
$$

Note that the one expectation in this equation is:

$$E[(Z^TY - Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1,\ldots,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]$$
$$= tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1,\ldots,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ_tG_tZ_t^TZ\}$$
$$+ \sigma_t^2 tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1,\ldots,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ\}$$
$$= [-\frac{K}{\sigma^2}\frac{1}{\xi^2}\frac{1}{\tau^2} - \frac{K}{\sigma^2}\frac{1}{\xi^2}(\frac{JK}{\sigma^2} + \frac{1}{\tau^2}) + 2(\frac{K}{\sigma^2})^2((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^2\xi)^{-1}$$
$$+ J(\frac{K}{\sigma^2})^4((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^2\xi^2)^{-1}](J^2K^2\tau_t^2 + T^2K^2\eta_t^2)$$
$$+ (\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-2}K^2(J\tau_t^2 + T\eta_t^2)$$
$$+ \sigma_t^2[-\frac{K}{\sigma^2}\frac{1}{\xi^2}\frac{1}{\tau^2}JK + JK(\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^{-2} - \frac{K}{\sigma^2}\frac{1}{\xi^2}(\frac{JK}{\sigma^2} + \frac{1}{\tau^2})JK$$
$$+ 2JK(\frac{K}{\sigma^2})^2((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^2\xi)^{-1} + J^2K(\frac{K}{\sigma^2})^4((\frac{K}{\sigma^2} + \frac{1}{\gamma^2})^2\xi^2)^{-1}]$$

3.6 Final system

Recall that the equation we are trying to solve is $E_{\alpha_t}[\sum_{m=1}^M \frac{\partial}{\partial\alpha}\log p_\alpha(Y_m|X_m)\Big|_{\alpha^*}] = \vec{0}$, and we have just found $E_{\alpha_t}\frac{\partial}{\partial\alpha}\log p_\alpha(Y_m|X_m)$ for a general cluster from sequence m (using notation from the main paper). Using the notation from the derivation above, the system is:

$$\sum_{m=1}^M E[\frac{\partial}{\partial\sigma^2}log(Pr(Y_m|X_m, Z_m))] = 0$$
$$\sum_{m=1}^M E[\frac{\partial}{\partial\tau^2}log(Pr(Y_m|X_m, Z_m))] = 0$$
$$\sum_{m=1}^M E[\frac{\partial}{\partial\gamma^2}log(Pr(Y_m|X_m, Z_m))] = 0$$

The only piece of these equations that depends on the sequence is T; from here on, we'll stop suppressing the index and call it T_m. Note that each of these equations can be written as a linear combination of T_m and/or T_m^2. So if we sum each equation over the sequences and divide by the number of sequences (M), we just need to replace T_m with $\frac{1}{M}\sum_{m=1}^M T_m$ and T_m^2 with $\frac{1}{M}\sum_{m=1}^M T_m^2$. In other words, the full system of equations is:

$$
\begin{aligned}
E[\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z))]\Big|_{T=\frac{1}{M}\sum_{m=1}^{M}T_m, T^2=\frac{1}{M}\sum_{m=1}^{M}T_m^2} &= 0 \\
E[\frac{\partial}{\partial\tau^2}log(Pr(Y|X,Z))]\Big|_{T=\frac{1}{M}\sum_{m=1}^{M}T_m, T^2=\frac{1}{M}\sum_{m=1}^{M}T_m^2} &= 0 \\
E[\frac{\partial}{\partial\gamma^2}log(Pr(Y|X,Z))]\Big|_{T=\frac{1}{M}\sum_{m=1}^{M}T_m, T^2=\frac{1}{M}\sum_{m=1}^{M}T_m^2} &= 0
\end{aligned}
$$

In this case, roots for this equation can be found directly using Mathematica; see supplemental Mathematica file for demonstration.

4 Treatment-fitted random time case

In this case, we are fitting a model with random treatment and intercept effects, but in truth the data comes from a model with random time and intercept effects.

4.1 Misspecified model

The misspecified model with random treatment effects is presented below. The fixed effects remain correctly specified (see Section 2).

$$
Y_i = X_i\theta + Z_i a_i + \epsilon_i
$$

$$
\begin{aligned}
Z_i = [\mathbf{1}_{JK}, \begin{matrix}\mathbf{0}_{K(J-T_i)} \\ \mathbf{1}_{KT_i}\end{matrix}], dim(JK,2) \\
a_i = [u_i, v_i]^T \sim MVN(0,G), dim(2,1) \\
G = diag(\tau^2, \eta^2), dim(2,2) \\
\epsilon_i \sim MVN(0, \sigma^2 I_{JK}), dim(JK,1)
\end{aligned}
$$

4.2 True model

The correctly specified model with random time effects is presented below. Note the 't' subscript, indicating that the matrix or parameter comes from the true model.

$$
Y_i = X_i\theta + Z_{it}a_{it} + \epsilon_{it}
$$

$$
\begin{aligned}
Z_{it} = [\mathbf{1}_{JK}, \begin{matrix}\mathbf{1}_K \\ \mathbf{0}_{JK-K}\end{matrix}, \ldots, \begin{matrix}\mathbf{0}_{JK-K} \\ \mathbf{1}_K\end{matrix}], dim(JK, J+1) \\
a_{it} = [u_{it}, w_{i1t}, \ldots, w_{iJt}]^T \sim MVN(0, G_t), dim(J+1, 1) \\
G_t = diag(\tau_t^2, \gamma_t^2, \ldots, \gamma_t^2), dim(J+1, J+1) \\
\epsilon_{it} \sim MVN(0, \sigma_t^2 I_{JK}), dim(JK, 1)
\end{aligned}
$$

4.3 Marginal misspecified likelihood

Recall that the equation we are trying to find roots for (Equation 3 in the main paper) is a sum over sequences, since the marginal likelihood is the same for every cluster within a sequence. For the next several sections, we'll be doing computations for a single sequence - that is, a single cluster within a sequence. To improve readability, the i subscripts in the notation presented above will be repressed.

Since $Y|a, X, Z \sim MVN(X\theta + Za, \sigma^2 I)$ under the mis-specified model, the conditional mis-specified likelihood is:

$$
\begin{aligned}
Pr(Y|a, X, Z) &= (2\pi)^{-JK/2}|\sigma^2 I|^{-1/2} exp\{-\frac{1}{2}(Y - (X\theta + Za))^T(\sigma^2 I)^{-1}(Y - (X\theta + Za))\} \\
&= (2\pi\sigma^2)^{-JK/2} exp\{-\frac{1}{2\sigma^2}(Y - (X\theta + Za))^T(Y - (X\theta + Za))\}
\end{aligned}
$$

Next, we integrate over the random effects to get the marginal mis-specified likelihood.

$$
\begin{aligned}
Pr(Y|X, Z) &= \int Pr(Y|a, X, Z) f(a|X, Z) da \qquad (4) \\
&= \int (2\pi\sigma^2)^{-JK/2} exp\{-\frac{1}{2\sigma^2}((Y - X\theta) - Za)^T((Y - X\theta) - Za)\} \\
&\quad (2\pi)^{-1}|G|^{-1/2} exp\{-\frac{1}{2}a^T G^{-1} a\} da \\
&= (2\pi\sigma^2)^{-JK/2}(2\pi)^{-1}(\tau^2\eta^2)^{-1/2} exp\{-\frac{1}{2\sigma^2}(Y - X\theta)^T(Y - X\theta)\} \\
&\quad \int exp\{-\frac{1}{2}(\frac{1}{\sigma^2}(-a^T Z^T(Y - X\theta) - (Y - X\theta)^T Za) + a^T(Z^T Z\frac{1}{\sigma^2} + G^{-1})a)\} da
\end{aligned}
$$

We wish to solve the integral by completing the square and obtaining a MVN($\frac{1}{\sigma^2}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta)$,$(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}$) pdf for a. Below are some useful facts about this covariance matrix.

$$
\begin{aligned}
Z^T Z\frac{1}{\sigma^2} + G^{-1} &= \frac{1}{\sigma^2}\begin{bmatrix} JK & TK \\ TK & TK \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau^2} & 0 \\ 0 & \frac{1}{\eta^2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{JK}{\sigma^2} + \frac{1}{\tau^2} & \frac{TK}{\sigma^2} \\ \frac{TK}{\sigma^2} & \frac{TK}{\sigma^2} + \frac{1}{\eta^2} \end{bmatrix}
\end{aligned}
$$

Since the determinant of the inverse is the inverse of the determinant, we can skip inverting for now (note that this matrix is invertible and symmetric). The determinant is:

$$
|Z^T Z\frac{1}{\sigma^2} + G^{-1}| = (\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2
$$

So, multiplying and dividing Equation 4 by the term below

$$
\begin{aligned}
&|(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}|^{-1/2} exp\{\frac{-1}{2}(\frac{1}{\sigma^2}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta))^T (Z^T Z\frac{1}{\sigma^2} + G^{-1}) \\
&\quad (\frac{1}{\sigma^2}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta))\}
\end{aligned}
$$

we get:

$$
\begin{aligned}
Pr(Y|X,Z) = {} & (2\pi\sigma^2)^{-JK/2}(\tau^2\eta^2)^{-1/2} exp\{-\frac{1}{2\sigma^2}(Y - X\theta)^T(Y - X\theta)\} \\
& ((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1/2} \\
& exp\{\frac{1}{2}(\frac{1}{\sigma^2})^2((Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta))^T(Z^T Y - Z^T X\theta)\} \\
= {} & (2\pi\sigma^2)^{-JK/2}(\tau^2\eta^2)^{-1/2}((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1/2} \\
& exp\{-\frac{1}{2\sigma^2}(Y - X\theta)^T(Y - X\theta) \\
& + \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^T Y - Z^T X\theta)^T(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T Y - Z^T X\theta)\}
\end{aligned}
$$

So the log marginal likelihood is:

$$
\begin{aligned}
log(Pr(Y|X,Z)) = &-\frac{JK}{2}log(2\pi\sigma^2) - \frac{1}{2}log(\tau^2\eta^2) \\
&- \frac{1}{2}log((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2) \\
&- \frac{1}{2\sigma^2}(Y - X\theta)^T(Y - X\theta) \\
&+ \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)
\end{aligned} \tag{5}
$$

4.4 Score equations

Fixed effects

Note that we are still keeping all the fixed effects in the vector θ, so this score equation is a vector also.

$$
\frac{\partial}{\partial\theta}log(Pr(Y|X,Z)) = -\frac{1}{\sigma^2}(-X^TY + X^TX\theta) - (\frac{1}{\sigma^2})^2X^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta)
$$

Residual variance

$$
\begin{aligned}
\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z)) = &-\frac{JK}{2}\frac{1}{\sigma^2} - \frac{1}{2}((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1} \\
&(\frac{-2JTK^2}{(\sigma^2)^3} - \frac{JK}{(\sigma^2)^2\eta^2} - \frac{KT}{(\sigma^2)^2\tau^2} + \frac{2K^2T^2}{(\sigma^2)^3}) \\
&+ \frac{1}{2(\sigma^2)^2}(Y - X\theta)^T(Y - X\theta) \\
&- (\frac{1}{\sigma^2})^3(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^TY - Z^TX\theta) \\
&+ \frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY - Z^TX\theta)^T[\frac{\partial}{\sigma^2}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)
\end{aligned}
$$

Note that $\frac{\partial}{\partial x}M(x)^{-1} = -M(x)^{-1}[\frac{\partial}{\partial x}M(x)]M(x)^{-1}$ and $\frac{\partial}{\sigma^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = Z^TZ\frac{-1}{(\sigma^2)^2}$, so

$$\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z)) = -\frac{JK}{2}\frac{1}{\sigma^2} - \frac{1}{2}((\frac{JK}{\sigma^2}+\frac{1}{\tau^2})(\frac{TK}{\sigma^2}+\frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1}$$
$$(\frac{-2JTK^2}{(\sigma^2)^3} - \frac{JK}{(\sigma^2)^2\eta^2} - \frac{KT}{(\sigma^2)^2\tau^2} + \frac{2K^2T^2}{(\sigma^2)^3})$$
$$+ \frac{1}{2(\sigma^2)^2}(Y-X\theta)^T(Y-X\theta)$$
$$-(\frac{1}{\sigma^2})^3(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta)$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(Z^TY-Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^TZ\frac{1}{(\sigma^2)^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)$$

Random effect variances

Similarly, because $\frac{\partial}{\tau^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = diag\{\frac{-1}{(\tau^2)^2}, 0\}$ and $\frac{\partial}{\eta^2}(Z^TZ\frac{1}{\sigma^2}+G^{-1}) = diag\{0, \frac{-1}{(\eta^2)^2}\}$,

$$\frac{\partial}{\partial\tau^2}log(Pr(Y|X,Z)) = -\frac{1}{2}\frac{1}{\tau^2} + \frac{1}{2}((\frac{JK}{\sigma^2}+\frac{1}{\tau^2})(\frac{TK}{\sigma^2}+\frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1}(\frac{1}{\tau^2})^2(\frac{KT}{\sigma^2}+\frac{1}{\eta^2})$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(\frac{1}{\tau^2})^2(Z^TY-Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{1,0\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)$$
$$\frac{\partial}{\partial\eta^2}log(Pr(Y|X,Z)) = -\frac{1}{2}\frac{1}{\eta^2} + \frac{1}{2}((\frac{JK}{\sigma^2}+\frac{1}{\tau^2})(\frac{TK}{\sigma^2}+\frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1}(\frac{1}{\eta^2})^2(\frac{JK}{\sigma^2}+\frac{1}{\tau^2})$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^2(\frac{1}{\eta^2})^2(Z^TY-Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}diag\{0,1\}(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)$$

4.5 Expectations

Next, we take expectations of the score equations, with respect to the true distribution of Y (that is, the correctly specified model). In this case (Normal outcomes, identity link), we know that θ is unbiased, so $E(Y|X,Z_t) = E(E(Y|X,Z_t,a_t)) = E(X\theta + Z_ta_t) = X\theta$. We will also make frequent use of the rule $E(x^TAx) = tr(Acov(x)) + E(x)^TAE(x)$.

Because it occurs so frequently, we will use $\xi = (\frac{KT}{\sigma^2}+\frac{1}{\eta^2})(\frac{JK}{\sigma^2}+\frac{1}{\tau^2}) - (\frac{KT}{\sigma^2})^2$ as shorthand; note that this is not the same definition of ξ used in the derivation of the time-fitted random treatment case.

Many of these expectations involve finding the trace of products of matrices. All the matrices involved have a convenient block structure, either 2-by-2 blocks or 3-by-3 blocks. Most of the matrix multiplication work is straightforward but lengthy, so is not described explicitly. However, to aid understanding some useful facts which are used many times in the calculations are presented below.

$$\begin{aligned}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1} &= \begin{bmatrix} \frac{JK}{\sigma^2} + \frac{1}{\tau^2} & \frac{TK}{\sigma^2} \\ \frac{TK}{\sigma^2} & \frac{TK}{\sigma^2} + \frac{1}{\eta^2} \end{bmatrix}^{-1} \\ &= \frac{1}{\xi}\begin{bmatrix} \frac{TK}{\sigma^2} + \frac{1}{\eta^2} & \frac{-TK}{\sigma^2} \\ \frac{-TK}{\sigma^2} & \frac{JK}{\sigma^2} + \frac{1}{\tau^2} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1} Z^T Z &= \frac{1}{\xi}\begin{bmatrix} \frac{TK}{\sigma^2} + \frac{1}{\eta^2} & \frac{-TK}{\sigma^2} \\ \frac{-TK}{\sigma^2} & \frac{JK}{\sigma^2} + \frac{1}{\tau^2} \end{bmatrix}\begin{bmatrix} JK & TK \\ TK & TK \end{bmatrix} \\ &= \frac{1}{\xi}\begin{bmatrix} (J-T)K\frac{TK}{\sigma^2} + JK\frac{1}{\eta^2} & TK\frac{1}{\eta^2} \\ TK\frac{1}{\tau^2} & (J-T)K\frac{TK}{\sigma^2} + TK\frac{1}{\tau^2} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}Z^T Z_t G_t Z_t^T Z &= \begin{bmatrix} JK & TK \\ K\mathbf{1}_{J-T} & \mathbf{0}_{J-T} \\ K\mathbf{1}_T & K\mathbf{1}_T \end{bmatrix}^T \begin{bmatrix} \tau_t^2 & \mathbf{0}_J^T \\ \mathbf{0}_J & \gamma_t^2 I_J \end{bmatrix} \begin{bmatrix} JK & TK \\ K\mathbf{1}_{J-T} & \mathbf{0}_{J-T} \\ K\mathbf{1}_T & K\mathbf{1}_T \end{bmatrix} \\ &= \begin{bmatrix} J^2K^2\tau_t^2 + JK^2\gamma_t^2 & JTK^2\tau_t^2 + TK^2\gamma_t^2 \\ JTK^2\tau_t^2 + TK^2\gamma_t^2 & T^2K^2\tau_t^2 + TK^2\gamma_t^2 \end{bmatrix}\end{aligned}$$

Fixed effects

$$\begin{aligned}E[\frac{\partial}{\partial\theta}log(Pr(Y|X,Z)] &= -\frac{1}{\sigma^2}(-X^T E[Y] + X^T X\theta) - (\frac{1}{\sigma^2})^2 X^T Z(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T E[Y] - Z^T X\theta) \\ &= -\frac{1}{\sigma^2}(-X^T X\theta + X^T X\theta) - (\frac{1}{\sigma^2})^2 X^T Z(Z^T Z\frac{1}{\sigma^2} + G^{-1})^{-1}(Z^T X\theta - Z^T X\theta) \\ &= \mathbf{0}\end{aligned}$$

Residual variance

$$E[\frac{\partial}{\partial\sigma^2}log(Pr(Y|X,Z))] = -\frac{JK}{2}\frac{1}{\sigma^2} - \frac{1}{2}((\frac{JK}{\sigma^2}+\frac{1}{\tau^2})(\frac{TK}{\sigma^2}+\frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1} \tag{6}$$
$$(\frac{-2JTK^2}{(\sigma^2)^3} - \frac{JK}{(\sigma^2)^2\eta^2} - \frac{KT}{(\sigma^2)^2\tau^2} + \frac{2K^2T^2}{(\sigma^2)^3})$$
$$+ \frac{1}{2(\sigma^2)^2}E[(Y-X\theta)^T(Y-X\theta)]$$
$$-(\frac{1}{\sigma^2})^3E[(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta)]$$
$$+\frac{1}{2}(\frac{1}{\sigma^2})^4E[(Z^TY-Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}](Z^TY-Z^TX\theta)]$$

To solve the first expectation in Equation 6, note that

$$\begin{aligned} E[(Y-X\theta)^T(Y-X\theta)] = tr\{cov(Y)\} &= tr\{Z_tG_tZ_t^T + \sigma_t^2 I_{JK}\} \\ &= JK(\tau_t^2+\gamma_t^2) + JK\sigma_t^2 \\ &= JK(\tau_t^2+\gamma_t^2+\sigma_t^2) \end{aligned}$$

To solve the second expectation in Equation 6, note that

$$E[(Z^TY - Z^TX\theta)^T(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}(Z^TY-Z^TX\theta)] = tr\{(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}cov(Z^TY)\}$$
$$\begin{aligned} &= tr\{(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^T(Z_tG_tZ_t^T+\sigma_t^2I_{JK})Z\} \\ &= tr\{(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^TZ_tG_tZ_t^TZ\} \\ &\quad + \sigma_t^2 tr\{(Z^TZ\frac{1}{\sigma^2}+G^{-1})^{-1}Z^TZ\} \\ &= \frac{1}{\xi}((\frac{TK}{\sigma^2}+\frac{1}{\eta^2})(J^2K^2\tau_t^2 + JK^2\gamma_t^2) + 2(\frac{-TK}{\sigma^2})(JTK^2\tau_t^2+TK^2\gamma_t^2) \\ &\quad + (\frac{JK}{\sigma^2}+\frac{1}{\tau^2})(T^2K^2\tau_t^2+TK^2\gamma_t^2)) \\ &\quad + \sigma_t^2\frac{1}{\xi}(2(J-T)K\frac{TK}{\sigma^2} + JK\frac{1}{\eta^2} + TK\frac{1}{\tau^2}) \end{aligned}$$

To solve the third expectation in Equation 6, note that

$$
\begin{aligned}
E[(Z^TY - Z^TX\theta)^T&[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)] \\
&= tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ_tG_tZ_t^TZ\} \\
&\quad + \sigma_t^2 tr\{(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}Z^TZ\} \\
&= \frac{1}{\xi^2}(((J-T)\frac{T^2K^3}{(\sigma^2)^2} + 2(J-T)K\frac{TK}{\sigma^2}\frac{1}{\eta^2} + JK\frac{1}{(\eta^2)^2})(J^2K^2\tau_t^2 + JK^2\gamma_t^2) \\
&\quad + 2(TK\frac{1}{\tau^2}\frac{1}{\eta^2} - (J-T)\frac{T^2K^3}{(\sigma^2)^2})(JTK^2\tau_t^2 + TK^2\gamma_t^2) \\
&\quad + ((J-T)\frac{JTK^3}{(\sigma^2)^2} + 2(J-T)K\frac{TK}{\sigma^2}\frac{1}{\tau^2} + TK\frac{1}{(\tau^2)^2})(T^2K^2\tau_t^2 + TK^2\gamma_t^2)) \\
&\quad + \sigma_t^2\frac{1}{\xi^2}(2(J-T)^2\frac{T^2K^4}{(\sigma^2)^2} + 2(J-T)JK^2\frac{TK}{\sigma^2}\frac{1}{\eta^2} + J^2K^2\frac{1}{(\eta^2)^2} \\
&\quad + 2T^2K^2\frac{1}{\tau^2}\frac{1}{\eta^2} + 2(J-T)K\frac{T^2K^2}{\sigma^2}\frac{1}{\tau^2} + T^2K^2\frac{1}{(\tau^2)^2})
\end{aligned}
$$

Random effect variances
We begin with the score equation for the random intercept effect.

$$
\begin{aligned}
E[\frac{\partial}{\partial\tau^2}log(Pr(Y|X,Z))] = &-\frac{1}{2}\frac{1}{\tau^2} + \frac{1}{2}((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1}(\frac{1}{\tau^2})^2(\frac{KT}{\sigma^2} + \frac{1}{\eta^2}) \\
&+ \frac{1}{2}(\frac{1}{\sigma^2})^2(\frac{1}{\tau^2})^2E[(Z^TY - Z^TX\theta)^T \\
&[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{1,0\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]
\end{aligned}
$$

Note that the one expectation in this equation is:

$$E[(Z^TY - Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{1,0\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]$$
$$= tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{1,0\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ_tG_tZ_t^TZ\}$$
$$+ \sigma_t^2 tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{1,0\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ\}$$
$$= \frac{1}{\xi^2}((\frac{TK}{\sigma^2} + \frac{1}{\eta^2})^2(J^2K^2\tau_t^2 + JK^2\gamma_t^2)$$
$$- 2(\frac{TK}{\sigma^2} + \frac{1}{\eta^2})\frac{TK}{\sigma^2}(JTK^2\tau_t^2 + TK^2\gamma_t^2) + (\frac{TK}{\sigma^2})^2(T^2K^2\tau_t^2 + TK^2\gamma_t^2))$$
$$+ \sigma_t^2\frac{1}{\xi^2}(JK(\frac{TK}{\sigma^2} + \frac{1}{\eta^2})^2 - (\frac{TK}{\sigma^2} + \frac{2}{\eta^2})\frac{T^2K^2}{\sigma^2})$$

The score equation for the random treatment effect has a similar form.

$$E[\frac{\partial}{\partial\eta^2}log(Pr(Y|X,Z))] = -\frac{1}{2}\frac{1}{\eta^2} + \frac{1}{2}((\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(\frac{TK}{\sigma^2} + \frac{1}{\eta^2}) - (\frac{TK}{\sigma^2})^2)^{-1}(\frac{1}{\eta^2})^2(\frac{JK}{\sigma^2} + \frac{1}{\tau^2})$$
$$+ \frac{1}{2}(\frac{1}{\sigma^2})^2(\frac{1}{\eta^2})^2E[(Z^TY - Z^TX\theta)^T$$
$$[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]$$

Note that the one expectation in this equation is:

$$E[(Z^TY - Z^TX\theta)^T[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}](Z^TY - Z^TX\theta)]$$
$$= tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ_tG_tZ_t^TZ\}$$
$$+ \sigma_t^2 tr\{[(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}diag\{0,1\}(Z^TZ\frac{1}{\sigma^2} + G^{-1})^{-1}]Z^TZ\}$$
$$= \frac{1}{\xi^2}((\frac{TK}{\sigma^2})^2(J^2K^2\tau_t^2 + JK^2\gamma_t^2) - 2\frac{TK}{\sigma^2}(\frac{JK}{\sigma^2} + \frac{1}{\tau^2})(JTK^2\tau_t^2 + TK^2\gamma_t^2)$$
$$+ (\frac{JK}{\sigma^2} + \frac{1}{\tau^2})^2(T^2K^2\tau_t^2 + TK^2\gamma_t^2)$$
$$+ \sigma_t^2(\frac{-T^2K^2}{\sigma^2}(\frac{JK}{\sigma^2} + \frac{2}{\tau^2}) + TK(\frac{JK}{\sigma^2} + \frac{1}{\tau^2})^2))$$

4.6 Final system

Recall that the equation we are trying to solve is $E_{\alpha_t}[\sum_{m=1}^{M}\frac{\partial}{\partial\alpha}\log p_\alpha(Y_m|X_m)\Big|_{\alpha^*}] = \vec{0}$, and we have just found $E_{\alpha_t}\frac{\partial}{\partial\alpha}\log p_\alpha(Y_m|X_m)$ for a general cluster from sequence m

(using notation from the main paper). Using the notation from the derivation above, the system is:

$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \sigma^2} log(Pr(Y_m|X_m, Z_m))] = 0$$
$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \tau^2} log(Pr(Y_m|X_m, Z_m))] = 0$$
$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \gamma^2} log(Pr(Y_m|X_m, Z_m))] = 0$$

The only piece of these equations that depends on the sequence is T; from here on, we'll stop suppressing the index and call it T_m. Unfortunately, it is not straightforward to write this system as a simple function of T_m's as we did in the time-fitted random treatment case. But since T_m is the only term that varies between sequences, the full system of equations can be written as:

$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \sigma^2} log(Pr(Y|X, Z))]\Big|_{T=T_m} = 0$$
$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \tau^2} log(Pr(Y|X, Z))]\Big|_{T=T_m} = 0$$
$$\sum_{m=1}^{M} E[\frac{\partial}{\partial \gamma^2} log(Pr(Y|X, Z))]\Big|_{T=T_m} = 0$$

Two roots exist which are not functions of T_m but cause each sequence-specific piece to be zero. Any other roots can be found numerically. See supplemental Mathematica file for closed-form roots and supplemental R file for numerical solutions.