
APPENDICES

In this chapter supplementary materials from different chapters are presented. Sections A.1–A.8 will provide materials related to Section ?? (structured horizontal splitting for CS). The materials related to Section ?? (structured horizontal splitting for AR1) will be given in Sections A.9–A.11. The Fieller’s method and delta method are explained in Section A.12. The combination rules for random vertical splitting for a CS case will be derived in Section A.13. In Section A.14, combination rules of multiple imputation for p -values are given. And finally, the surrogate model we have used in thesis is presented in Section A.15.

A.1 Incompleteness in the compound-symmetry model

Based on (??) and the multiplicity of the cluster sizes, the sufficient statistics are:

$$W_{1k} = \sum_{i=1}^{c_k} \sum_{j=1}^{n_k} Y_{ij}^{(k)}, \quad (\text{A.1.1})$$

$$W_2 = \sum_{k=1}^L \sum_{i=1}^{c_k} \sum_{j=1}^{n_k} \left(Y_{ij}^{(k)} \right)^2, \quad (\text{A.1.2})$$

$$W_{3k} = \sum_{i=1}^{c_k} \left(\sum_{j=1}^{n_k} Y_{ij}^{(k)} \right)^2, \quad (\text{A.1.3})$$

$$W_{4k} = c_k. \quad (\text{A.1.4})$$

The conditional and marginal expectations of (A.1.1)–(A.1.4) are:

$$E(W_{1k}|c_k) = c_k n_k \mu,$$

$$E(W_{1k}) = N \mu \pi_k n_k,$$

$$E(W_2|c_k) = \sum_{k=1}^L c_k n_k (\sigma^2 + d + \mu^2),$$

$$E(W_2) = N (\sigma^2 + d + \mu^2) \sum_{k=1}^L \pi_k n_k,$$

$$E(W_{3k}|c_k) = c_k \left\{ n_k (\sigma^2 + d + \mu^2) + n_k (n_k - 1) (d + \mu^2) \right\},$$

$$E(W_{3k}) = N \pi_k n_k \left\{ (\sigma^2 + d + \mu^2) + (n_k - 1) (d + \mu^2) \right\},$$

$$E(W_{4k}) = N \pi_k.$$

Group all sufficient statistics in \mathbf{W} and define a function

$$g(\mathbf{w}) = \sum_{k=1}^L \lambda_k \frac{w_{1k}}{w_{4k}}. \quad (\text{A.1.5})$$

Then,

$$E \{g(\mathbf{W}|\mathbf{W}_4)\} = \sum_{k=1}^L \lambda_k \frac{E(W_{1k}|W_{4k})}{W_{4k}} = \mu \sum_{k=1}^L \lambda_k n_k,$$

and hence

$$E \{g(\mathbf{W})\} = \mu \sum_{k=1}^K \lambda_k n_k.$$

Thus, every solution $\boldsymbol{\lambda} \perp \mathbf{n}$, where $\mathbf{n} = (n_1, \dots, n_K)'$, provides a counterexample, establishing incompleteness.

Such a vector $\boldsymbol{\lambda}$ exists if and only if $K \geq 2$, for which it is assumed that at least two $c_k > 0$ (i.e., at least two different cluster sizes occur).

A.2 Likelihood-based estimation of the CS model

A.2.1 Score functions

The score function has components:

$$\frac{\partial \ell}{\partial \mu_k} = \frac{1}{\sigma_k^2 + n_k d_k} \left(\sum_{i=1}^{c_k} \sum_{j=1}^{n_1} y_{ij}^{(k)} - c_k n_k \mu_k \right), \quad (\text{A.2.1})$$

$$\frac{\partial \ell}{\partial \sigma_k^2} = \frac{-c_k n_k}{2\sigma_k^2} \cdot \frac{\sigma_k^2 + (n_k - 1)d_k}{\sigma_k^2 + n_k d_k} + \frac{c_k n_k S_k}{2\sigma_k^4} \quad (\text{A.2.2})$$

$$- \frac{d_k(2\sigma_k^2 + n_k d_k)c_k n_k T_k}{2\sigma_k^4(\sigma_k^2 + n_k d_k)^2}, \quad (\text{A.2.3})$$

$$\frac{\partial \ell}{\partial d_k} = \frac{-c_k n_k}{2(\sigma_k^2 + n_k d_k)} + \frac{c_k n_k T_k}{2(\sigma_k^2 + n_k d_k)^2}, \quad (\text{A.2.4})$$

with

$$S_k = \frac{1}{c_k n_k} Q_k = \frac{1}{c_k n_k} \sum_{i=1}^{c_k} \mathbf{Z}_i^{(k)'} \mathbf{Z}_i^{(k)}, \quad (\text{A.2.5})$$

$$T_k = \frac{1}{c_k n_k} R_k = \frac{1}{c_k n_k} \sum_{i=1}^{c_k} \mathbf{Z}_i^{(k)'} \mathbf{J}_{\mathbf{n}_k} \mathbf{Z}_i^{(k)}. \quad (\text{A.2.6})$$

A.2.2 Lack of closed-form solution when $K \geq 2$

The lack of a closed form when $K \geq 2$ is well known, but we highlight a few relevant features here. More detail is given in Supplementary Materials A.3. Function (4.4) can be turned into the log-likelihood kernel for the conventional situation where there is a common mean

parameter and common variance components across all cluster sizes, i.e., $\ell(\mu, \sigma^2, d)$. The score functions follow from summing the terms in (A.2.1)–(A.2.4) across cluster sizes:

$$\begin{aligned}\frac{\partial \ell}{\partial \mu} &= \sum_{k=1}^K \frac{\partial \ell}{\partial \mu_k} \Big|_{\mu_k=\mu}, \\ \frac{\partial \ell}{\partial \sigma^2} &= \sum_{k=1}^K \frac{\partial \ell}{\partial \sigma_k^2} \Big|_{\sigma_k^2=\sigma^2}, \\ \frac{\partial \ell}{\partial d} &= \sum_{k=1}^K \frac{\partial \ell}{\partial d_k} \Big|_{d_k=d}.\end{aligned}\tag{A.2.7}$$

Solving the score equation in (A.2.7) for the mean, using that

$$\Sigma_{n_k}^{-1} = \frac{1}{\sigma^2} I_{n_k} - \frac{d}{\sigma^2(\sigma^2 + n_k d)} J_{n_k},$$

leads to the identity:

$$\hat{\mu} = \frac{\sum_{k=1}^K \frac{n_k c_k}{\sigma^2 + n_k d} \bar{Y}^{(k)}}{\sum_{k=1}^K \frac{n_k c_k}{\sigma^2 + n_k d}} = \frac{\sum_{k=1}^K \frac{n_k c_k}{\sigma^2 + n_k d} \widehat{\mu}_k}{\sum_{k=1}^K \frac{n_k c_k}{\sigma^2 + n_k d}},\tag{A.2.8}$$

where $\widehat{\mu}_k$ as in (4.5). For the variance components, only implicit identities follow; they are functions of (A.2.5)–(A.2.6). These take the form of high-degree polynomials, for which no general explicit solution exists. While (A.2.8) is explicit, it is a weighted average of the cluster-size specific averages $\bar{Y}^{(k)}$, with weights depending on the variance components. This, combined with the result for the variance components, implies that there is no explicit solution, unless the variance components are known or the

cluster size is constant.

A.3 Full likelihood

Referring to the conventional situations, i.e. $\ell(\mu, \sigma^2, d)$ in (4.2) and the score equation in (A.2.7), also second derivatives can be calculated:

$$\frac{\partial^2 \ell}{\partial \mu^2} = \sum_{k=1}^K \frac{-c_k n_k}{\sigma^2 + n_k d} \quad (\text{A.3.1})$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \mu} = \sum_{k=1}^K \frac{-1}{(\sigma^2 + n_k d)^2} \left(\sum_{i=1}^{c_k} \sum_{j=1}^{n_k} y_{ij}^{(k)} - c_k n_k \mu_k \right) \quad (\text{A.3.2})$$

$$\frac{\partial^2 \ell}{\partial d \partial \mu_k} = \sum_{k=1}^K \frac{-n_k}{(\sigma^2 + n_k d)^2} \left(\sum_{i=1}^{c_k} \sum_{j=1}^{n_k} y_{ij}^{(k)} - c_k n_k \mu_k \right) \quad (\text{A.3.3})$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} = \sum_{k=1}^K \left(\frac{-1}{\sigma^4} + \frac{d(2\sigma^2 + n_k d)n_k}{\sigma^4(\sigma^2 + n_k d)^2} \right) \sum_{i=1}^{c_k} \sum_{j=1}^{n_k} Z_{ij}^{(k)} \quad (\text{A.3.4})$$

$$\begin{aligned} \frac{\partial^2 \ell}{(\partial \sigma^2)^2} = \sum_{k=1}^K \left(\frac{c_k n_k}{2\sigma^4} \cdot \frac{\sigma^2 + (n_k - 1)d}{\sigma^2 + n_k d} - \frac{c_k n_k}{2\sigma^2} \cdot \frac{d}{(\sigma^2 + n_k d)^2} - \frac{c_k n_k S_k}{\sigma^6} \right. \\ \left. - d c_k n_k T_k \frac{\sigma^2(\sigma^2 + n_k d) - (2\sigma^2 + n_k d)^2}{\sigma^6(\sigma^2 + n_k d)^3} \right) \quad (\text{A.3.5}) \end{aligned}$$

$$\frac{\partial^2 \ell}{\partial d \partial \sigma^2} = \sum_{i=1}^K \left(\frac{c_k n_k}{2(\sigma^2 + n_k d)} - \frac{c_k n_k T_k}{\sigma^4} \cdot \frac{(\sigma^2 + n_k d)^2 - n_k d(2\sigma^2 + n_k d)}{(\sigma^2 + n_k d)^3} \right) \quad (\text{A.3.6})$$

$$\frac{\partial^2 \ell}{\partial \mu \partial d} = \sum_{k=1}^K \frac{-n_k}{(\sigma^2 + n_k d)^2} \sum_{i=1}^{c_k} \sum_{j=1}^{n_k} Z_{ij}^{(k)} \quad (\text{A.3.7})$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial d} = \sum_{k=1}^K \left(\frac{c_k n_k}{2(\sigma^2 + n_k d)^2} - \frac{c_k n_k T_k}{(\sigma^2 + n_k d)^3} \right) \quad (\text{A.3.8})$$

$$\frac{\partial^2 \ell}{\partial d^2} = \sum_{k=1}^K \left(\frac{c_k n_k^2}{2(\sigma^2 + n_k d)^2} - \frac{c_k n_k^2 T_k}{(\sigma^2 + n_k d)^3} \right) \quad (\text{A.3.9})$$

Should we use conditional likelihood, then the log-likelihood's kernel equals:

$$\begin{aligned}
 L \propto & \prod_{i=1}^k \frac{1}{(2\pi)^{n_1/2} |\Sigma_{n_1}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_{i1} - \boldsymbol{\mu}_{n_1})' \Sigma_{n_1}^{-1} (\mathbf{y}_{i1} - \boldsymbol{\mu}_{n_1}) \right\} \\
 & \times \left[\frac{\Phi(\alpha + \mathbf{y}'_{i1} \boldsymbol{\beta})}{\Phi \left(\frac{\alpha + \boldsymbol{\mu}'_{n_1} \boldsymbol{\beta}}{\sqrt{1 + \boldsymbol{\beta}' \Sigma_{n_1} \boldsymbol{\beta} / n_1}} \right)} \right] \\
 & \times \prod_{i=k+1}^N \frac{1}{(2\pi)^{n_2/2} |\Sigma_{n_2}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_{n_2})' \Sigma_{n_2}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_{n_2}) \right\} \\
 & \times \left[\frac{1 - \Phi(\alpha + \mathbf{y}'_{i1} \boldsymbol{\beta})}{1 - \Phi \left(\frac{\alpha + \boldsymbol{\mu}'_{n_1} \boldsymbol{\beta}}{\sqrt{1 + \boldsymbol{\beta}' \Sigma_{n_1} \boldsymbol{\beta} / n_1}} \right)} \right] \quad (\text{A.3.10})
 \end{aligned}$$

$$\begin{aligned}
 \ell \propto & -\frac{1}{2} \sum_{i=1}^k \left\{ \ln |\Sigma_{n_1}| + (\mathbf{y}_i - \boldsymbol{\mu}_{n_1})' \Sigma_{n_1}^{-1} (\mathbf{y}_{i1} - \boldsymbol{\mu}_{n_1}) \right\} \\
 & - k \ln \Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \\
 & - \frac{1}{2} \sum_{i=k+1}^N \left\{ \ln |\Sigma_{n_2}| + (\mathbf{y}_i - \boldsymbol{\mu}_{n_2})' \Sigma_{n_2}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_{n_2}) \right\} \\
 & - (N - k) \ln \left\{ 1 - \Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \right\}, \quad (\text{A.3.11})
 \end{aligned}$$

with $\tilde{\alpha} = \frac{\alpha}{\sqrt{1+\boldsymbol{\beta}'\Sigma_{n_1}\boldsymbol{\beta}/n_1}}$ and $\tilde{\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}}{\sqrt{1+\boldsymbol{\beta}'\Sigma_{n_1}\boldsymbol{\beta}/n_1}}$. The corresponding score equations are:

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} = & \frac{1}{\sigma^2 + n_1 d} \left(\sum_{i=1}^k \sum_{j=1}^{n_1} \mathbf{y}_{ij} - kn_1 \mu \right) - kn_1 \mathbf{j}'_{n_1} \tilde{\boldsymbol{\beta}} \frac{\phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})}{\Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})} \\ & + \frac{1}{\sigma^2 + n_2 d} \left(\sum_{i=k+1}^N \sum_{j=1}^{n_2} \mathbf{y}_{ij} - (N-k)n_2 \mu \right) \\ & - (N-k)n_2 \mathbf{j}'_{n_1} \tilde{\boldsymbol{\beta}} \frac{\phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})}{\Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})}, \end{aligned} \quad (\text{A.3.12})$$

with $\frac{\partial \ell}{\partial \sigma^2}$ and $\frac{\partial \ell}{\partial d}$ identical to (A.2.3) and (A.2.4). The components of the Hessian are:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \mu^2} = & \frac{-kn_1}{\sigma^2 + n_1 d} - \frac{(N-k)n_2}{\sigma^2 + n_2 d} \\ & - kn_1 \mathbf{j}'_{n_1} \tilde{\boldsymbol{\beta}} \left[\frac{-\Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot \phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot (\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot \mathbf{j}_{n_1} \tilde{\boldsymbol{\beta}}}{\Phi^2(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})} \right. \\ & \quad \left. - \frac{\phi^2(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot \mathbf{j}_{n_1} \tilde{\boldsymbol{\beta}}}{\Phi^2(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})} \right] \\ & \quad + (N-k)n_2 \mathbf{j}'_{n_1} \tilde{\boldsymbol{\beta}} \\ & \times \left[\frac{-(1 - \Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}})) \cdot \phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot (\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot \mathbf{j}_{n_1} \tilde{\boldsymbol{\beta}}}{(1 - \Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}))^2} \right. \\ & \quad \left. + \frac{\phi^2(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}) \cdot \mathbf{j}_{n_1} \tilde{\boldsymbol{\beta}}}{(1 - \Phi(\tilde{\alpha} + \boldsymbol{\mu}'_{n_1} \tilde{\boldsymbol{\beta}}))^2} \right] \end{aligned} \quad (\text{A.3.13})$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \sigma^2 \partial \mu} = & \frac{-1}{(\sigma^2 + n_1 d)^2} \left(\sum_{i=1}^k \sum_{j=1}^{n_1} \mathbf{y}_{ij} - k n_1 \mu \right) \\ & - \frac{1}{(\sigma^2 + n_2 d)^2} \left(\sum_{i=k+1}^N \sum_{j=1}^{n_2} \mathbf{y}_{ij} - (N - k) n_2 \mu \right) \end{aligned} \quad (\text{A.3.14})$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial d \partial \mu} = & \frac{-n_1}{(\sigma^2 + n_1 d)^2} \left(\sum_{i=1}^k \sum_{j=1}^{n_1} \mathbf{y}_{ij} - k n_1 \mu \right) \\ & - \frac{n_2}{(\sigma^2 + n_2 d)^2} \left(\sum_{i=k+1}^N \sum_{j=1}^{n_2} \mathbf{y}_{ij} - (N - k) n_2 \mu \right) \end{aligned} \quad (\text{A.3.15})$$

with $\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2}$, $\frac{\partial^2 \ell}{(\partial \sigma^2)^2}$, $\frac{\partial^2 \ell}{\partial d \partial \sigma^2}$, $\frac{\partial^2 \ell}{\partial \mu \partial d}$, $\frac{\partial^2 \ell}{\partial \sigma^2 \partial d}$, and $\frac{\partial^2 \ell}{\partial (d)^2}$ identical to (A.3.4), (A.3.5), (A.3.6), (A.3.7), (A.3.8), and (A.3.9).

A.4 Pseudo-likelihood for split samples

A.4.1 General considerations

Informally, a pseudo-likelihood function is one that replaces a computationally intractable or slow-to-maximize likelihood function, with another function that still produces a consistent and asymptotically normal estimator when maximised, i.e., by setting the first derivatives of the log pseudo-likelihood equal to zero and solving the resultant pseudo score equations. An early reference is Arnold and Strauss (1991), and details on various forms of pseudo-likelihood can be found in (Molenberghs and Verbeke, 2005, Ch. 9, 12, 21, 22, 24, and 25). In the current

clustered setting, the likelihood contribution of a cluster is often replaced by a product of contributions for various sub-vectors. In some cases, such a sub-vector can be conditioned upon another sub-vector.

Fieuws and Verbeke (2006) and Fieuws et al. (2006) used pseudo-likelihood to fit mixed models to high-dimensional multivariate longitudinal data. They supplemented the standard method with an additional device. They first replaced a set of M longitudinal sequences by the $M(M-1)/2$ longitudinal pairs. This in itself is a standard application of pseudo-likelihood. They then assumed that each pair has its own parameter vector. Symbolically, this can be written as:

$$p\ell(\boldsymbol{\theta}) \equiv p\ell(\mathbf{y}_{1i}, \mathbf{y}_{2i}, \dots, \mathbf{y}_{Mi} | \boldsymbol{\theta}) = \sum_{r < s} \ell(\mathbf{y}_{ri}, \mathbf{y}_{si} | \boldsymbol{\theta}_{rs}), \quad (\text{A.4.1})$$

where \mathbf{Y}_{ri} is the r th sequence for subject i . In (A.4.1), $\boldsymbol{\theta}$ results from stacking all $M(M-1)/2$ pair-specific parameter vectors $\boldsymbol{\theta}_{rs}$. The actual parameter vector of interest is $\boldsymbol{\theta}^*$, the set of non-redundant parameters is $\boldsymbol{\theta}$.

To obtain $\boldsymbol{\theta}^*$, Fieuws and Verbeke (2006) take averages of all available estimates for that specific parameter, implying that $\hat{\boldsymbol{\theta}}^* = A\hat{\boldsymbol{\theta}}$ for an appropriate linear combination matrix A . Further, combining this step with general pseudo-likelihood inference, a sandwich estimator is used:

$$\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*) = \sqrt{N}(A\hat{\boldsymbol{\theta}} - A \quad (\text{A.4.2})$$

$$bm\boldsymbol{\theta}) \overset{\text{approx.}}{\approx} N(\mathbf{0}, AI_0^{-1}I_1I_0^{-1}A'), \quad (\text{A.4.3})$$

where

$$I_0(\boldsymbol{\theta}) = E \left[\frac{\partial^2 p\ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \right], \quad I_1(\boldsymbol{\theta}) = E \left[\left(\frac{\partial p\ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \cdot \frac{\partial p\ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]. \quad (\text{A.4.4})$$

Molenberghs et al. (2011) took a very similar route to partition a potentially large sample into sub-samples.

To fix ideas, consider log-likelihood (4.4). When used as an instrument to estimate a single vector (μ, σ^2, d) , this function can be viewed a pseudo-likelihood. This setting can be generalized by assuming that a dataset, consisting of repeated measures per subject, is divided into K subgroups, each containing c_k independent replicates. Consider the pseudo-likelihood:

$$p\ell(\boldsymbol{\theta}) = \sum_{k=1}^K \ell(\boldsymbol{\theta}_k | \mathbf{y}_1^{(k)}, \dots, \mathbf{y}_{c_k}^{(k)}). \quad (\text{A.4.5})$$

While the underlying principle is similar to (A.4.1), it is not identical. The similarities are: (1) all $\boldsymbol{\theta}_k$ are assumed to be different, allowing for separate, even parallel, estimation; (2) $\boldsymbol{\theta}$ stacks all vectors $\boldsymbol{\theta}_k$; (3) the parameter of interest $\boldsymbol{\theta}^*$, is found from an appropriate combination of the $\boldsymbol{\theta}_k$. Parallel estimation was also followed by Scott et al. (2016) and Neiswanger et al. (2013).

There are important differences, however. Here, and in the remainder of the article, we assume that $\ell(\boldsymbol{\theta}_k | \mathbf{y}_1^{(k)}, \dots, \mathbf{y}_{c_k}^{(k)})$ is the likelihood that we would have, should group k be the only one in the data. That is,

the individual likelihood contributions are not altered, rather the data are partitioned. This is similar to the independent partitioning done by Molenberghs et al. (2011). In line with their derivations, (A.4.3) can also be used here. Given that $\ell_k(\boldsymbol{\theta}_k)$ is a genuine likelihood, its contributions to $I_0(\boldsymbol{\theta})$ and $I_1(\boldsymbol{\theta})$ are identical, up to the sign. As a result, $I_0(\boldsymbol{\theta})^{-1}I_1(\boldsymbol{\theta})I_0(\boldsymbol{\theta})^{-1} = -I_0(\boldsymbol{\theta})^{-1}$, a block-diagonal matrix with blocks of the form $I_0(\boldsymbol{\theta}_k)$. We now turn to the split-sample case.

A.4.2 Pseudo-likelihood for split sample

Molenberghs et al. (2011) chose

$$A = \frac{1}{K}(I, \dots, I) \quad (\text{A.4.6})$$

to pass from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$. This is a sensible choice in the i.i.d. setting (e.g., when all clusters in a CS model have the same size) and with the same number of subjects per sub-sample. The estimator and precision estimator then become:

$$\hat{\boldsymbol{\theta}}^* = \frac{1}{K} \sum_{k=1}^K \hat{\boldsymbol{\theta}}_k, \quad (\text{A.4.7})$$

$$\text{var}(\hat{\boldsymbol{\theta}}^*) = \frac{1}{K} H_{\hat{\boldsymbol{\theta}}}^{-1}, \quad (\text{A.4.8})$$

with $H_{\hat{\boldsymbol{\theta}}}^{-1} = -I_0(\boldsymbol{\theta}_k)$. In this special case, the expected information matrices are identical. Alternatively, one can use the observed information

matrices, and then use instead:

$$\frac{1}{K^2} \sum_{k=1}^K \mathcal{H}_{\hat{\theta},k}^{-1}. \quad (\text{A.4.9})$$

where $\mathcal{H}_{\hat{\theta},k}$ is the observed information for sub-sample k .

In this particular case, pseudo-likelihood produces the same estimator as full likelihood. This stems from the fact that all subjects follow the same distribution, in contrast to, for example, the setting set out at the start of Section 4. Should the subjects have identical distributions, but with c_k variables, the above results can be modified accordingly. For example, (A.4.7) would be replaced by

$$\hat{\boldsymbol{\theta}}^* = \sum_{k=1}^K \frac{c_k}{N} \hat{\boldsymbol{\theta}}_k,$$

with similar modification to precision estimation. In this case, full likelihood would be obtained.

In the next section, we will consider the more general case where different subjects may have a different distribution such as, for example, the CS case with a different number of measurements per cluster.

A.5 Derivation of optimal scalar weights for compound-symmetry case

To find the optimal scalar weight with minimum variance for μ we use the method of Lagrange with the constraint that the weights a_k need to sum to 1:

$$Q = \sum_{k=1}^K a_k^2 \frac{\sigma^2 + n_k d}{c_k n_k} - \lambda \left(\sum_{k=1}^K a_k - 1 \right). \quad (\text{A.5.1})$$

Solving the first partial derivative we become an expression for a_k involving λ . Summing this one produces an expression for λ , leading to the complete formula for a_k . Precisely, the system of equations is:

$$\frac{\partial Q}{\partial a_k} = 2a_k \frac{\sigma^2 + n_k d}{c_k n_k} - \lambda = 0, \quad (\text{A.5.2})$$

$$\frac{\partial Q}{\partial \lambda} = \sum_{k=1}^K a_k - 1 = 0, \quad (\text{A.5.3})$$

or, alternatively:

$$a_k = \frac{\lambda}{2} \frac{c_k n_k}{\sigma^2 + n_k d}, \quad (\text{A.5.4})$$

$$\lambda = \left(\frac{1}{2} \sum_{k=1}^K \frac{c_k n_k}{\sigma^2 + n_k d} \right)^{-1}, \quad (\text{A.5.5})$$

and hence

$$a_k = \frac{\frac{c_k n_k}{\sigma^2 + n_k d}}{\sum_{m=1}^K \frac{c_m n_m}{\sigma^2 + n_m d}}.$$

In the same manner, expressions for b_k and g_k can be found, as we will show next. For σ^2 :

$$Q = 2\sigma^4 \sum_{k=1}^K b_k^2 \frac{1}{c_k(n_k - 1)} - \lambda \left(\sum_{k=1}^K b_k - 1 \right),$$

producing the system:

$$\frac{\partial Q}{\partial b_k} = \frac{4\sigma^4 b_k}{c_k(n_k - 1)} - \lambda = 0, \quad (\text{A.5.6})$$

$$\frac{\partial Q}{\partial \lambda} = \sum_{k=1}^K b_k - 1 = 0, \quad (\text{A.5.7})$$

which can be rewritten as:

$$4\sigma^4 b_k = \lambda c_k(n_k - 1), \quad (\text{A.5.8})$$

$$\lambda = \frac{4\sigma^4}{\sum_{k=1}^K c_k(n_k - 1)}, \quad (\text{A.5.9})$$

and finally:

$$b_k = \frac{c_k(n_k - 1)}{\sum_{m=1}^K c_m(n_m - 1)}.$$

For d , the objective function is:

$$Q = \sum_{k=1}^K g_k^2 v_k - \lambda \left(\sum_{k=1}^K g_k - 1 \right),$$

with

$$v_k = \frac{2}{c_k n_k} \left(\frac{\sigma^4}{n_k - 1} + 2d\sigma^2 + n_k d^2 \right).$$

The system of equations now is:

$$\frac{\partial Q}{\partial g_k} = 2g_k v_k - \lambda = 0, \quad (\text{A.5.10})$$

$$\frac{\partial Q}{\partial \lambda} = \sum_{k=1}^K g_k - 1 = 0, \quad (\text{A.5.11})$$

leading to:

$$g_k = \lambda \frac{1}{2v_k}, \quad (\text{A.5.12})$$

$$\lambda = \frac{2}{\sum_{k=1}^K \frac{1}{v_k}}, \quad (\text{A.5.13})$$

giving the solution:

$$g_k = \frac{\frac{c_k n_k}{\frac{\sigma^4}{n_k - 1} + 2d\sigma^2 + n_k d^2}}{\sum_{m=1}^K \frac{c_m n_m}{\frac{\sigma^4}{n_m - 1} + 2d\sigma^2 + n_m d^2}}.$$

A.5.1 Cluster-by-cluster analysis

We study the case of the most extreme partitioning, i.e., where each of the clusters is analyzed separately. This can be relevant in cases with perhaps a limited number of very to extremely large clusters. This means that $c_k \equiv 1$ throughout. Clearly, the n_k will then no longer be unique. We will examine this case in detail, and contrast a first weighted estimator with an *ad hoc* one.

The Weighted Estimator for the Cluster-by-cluster Case

The estimator follows from setting $c_k \equiv 1$ and hence $K \equiv N$ throughout. For example, this special case can easily be considered for all expressions in Sections 6.1.1-6.1.3. Because c_k enters the inverse of the variance-covariance matrix multiplicatively, as is seen from (4.8)-(4.9), the optimal estimator that is obtained when each cluster is considered to be its own stratum, is identical to the one obtained when strata are defined in terms of all clusters of a given size. The same is true for the scalar weights.

It is insightful to consider in more detail the special case where further the cluster sizes are all identical to n . One then easily obtains:

$$\hat{\mu} = \frac{1}{Nn} \sum_{i=1}^N \sum_{j=1}^n Y_{ij}, \quad (\text{A.5.14})$$

$$\hat{\sigma}^2 = \frac{1}{Nn(n-1)} \left(n \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i - \sum_{i=1}^N \mathbf{Z}'_i J_n \mathbf{Z}_i \right) \quad (\text{A.5.15})$$

$$= \frac{1}{Nn(n-1)} (nQ - R), \quad (\text{A.5.16})$$

$$\hat{d} = \frac{1}{Nn(n-1)} \left(\sum_{i=1}^N \mathbf{Z}'_i J_n \mathbf{Z}_i - \sum_{i=1}^N \mathbf{Z}'_i \mathbf{Z}_i \right) \quad (\text{A.5.17})$$

$$= \frac{1}{Nn(n-1)} (Q - R), \quad (\text{A.5.18})$$

with obvious notation for Q and R , inspired by (A.2.5)–(A.2.6). The corresponding variance-covariance elements, similar in spirit to (4.8)–

(4.9), are:

$$\text{var}(\hat{\mu}) = \frac{\sigma^2 + nd}{Nn}, \quad (\text{A.5.19})$$

$$\text{var} \begin{pmatrix} \hat{\sigma}^2 \\ \hat{d} \end{pmatrix} = \begin{pmatrix} \frac{2\sigma^4}{N(n-1)} & -\frac{2\sigma^4}{Nn(n-1)} \\ -\frac{2\sigma^4}{Nn(n-1)} & \frac{2}{Nn} \left[\frac{\sigma^4}{n-1} + 2\sigma^2 d + nd^2 \right] \end{pmatrix}. \quad (\text{A.5.20})$$

This estimator coincides with the MLE, as is known from Molenberghs et al. (2011).

A Two-stage Estimator for Compound Symmetry

In linear mixed models, there is a method of estimation, sometimes called the two-stage approach (Laird et al., 1982; Verbeke and Molenberghs, 2000), in which each cluster is analyzed separately to begin with, using linear regression, after which the cluster-specific parameters are summarized into fixed effects. Although the above cluster-by-cluster analysis is superficially similar to this, it is not equivalent. In particular, there is no bias (as can be seen in the two stage method), and the maximum likelihood estimator is recovered.

This approach is most useful when cluster sizes are not constant, and in models that are more complex than compound symmetry. However, to gain some insight, we develop the details of the method for the CS model with constant cluster size.

For the mean, (A.5.14) is retained, as the average of the cluster-specific

averages \bar{Y}_i . Further, define:

$$s^2 = \frac{1}{Nn} \sum_{i=1}^N \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2, \quad (\text{A.5.21})$$

$$t^2 = \frac{1}{N} \sum_{i=1}^N (\bar{Y}_i - \hat{\mu})^2. \quad (\text{A.5.22})$$

Straightforward algebra shows:

$$E(s^2) = \frac{n-1}{n} \sigma^2, \quad (\text{A.5.23})$$

$$E(t^2) = \frac{N-1}{N} \left(d + \frac{1}{n} \sigma^2 \right). \quad (\text{A.5.24})$$

Should n and N approach infinity, then it follows that s^2 and t^2 are asymptotically unbiased estimators for σ^2 and d , respectively. However, this is not always reasonable. In applications such as the NTP data (Section 2), it is fair to say that the cluster size has a biological upper limit. In other situations, however, such as meta-analyses, it is sensible to assume that both n and N approach infinity.

In the next section, we will study the consequences of removing the bias. For now, a small, obvious modification is:

$$s_*^2 = \frac{n}{n-1} s^2, \quad (\text{A.5.25})$$

$$t_*^2 = \frac{N}{N-1} t^2. \quad (\text{A.5.26})$$

Now, s_*^2 is unbiased, while $E(t_*^2) = d + \sigma^2/n$, the bias σ^2/n can be made to disappear asymptotically provided it is sensible to let n grow large.

It is of interest to consider the variance-covariance structure of the estimators s^2 , t^2 , s_*^2 , and t_*^2 , as well as to make relative efficiency considerations. This will be done next.

Connections Between Estimators

Comparing algebraic expressions (A.5.16)–(A.5.18) with (A.5.21)–(A.5.22), leads to the linear relationships:

$$s^2 = \frac{n-1}{n} \widehat{\sigma^2} + 0 \cdot \widehat{d}, \quad (\text{A.5.27})$$

$$t^2 = \frac{N-1}{Nn} \widehat{\sigma^2} + \frac{N-1}{N} \widehat{d}. \quad (\text{A.5.28})$$

Relationships (A.5.27)–(A.5.28) can be combined with (A.5.20) to produce:

$$\text{var} \begin{pmatrix} s^2 \\ t^2 \end{pmatrix} = \begin{pmatrix} \frac{2(n-1)\sigma^4}{Nn^2} & 0 \\ 0 & \frac{2(N-1)^2}{N^2n} \left[\frac{\sigma^4}{n} + 2\sigma^2d + nd^2 \right] \end{pmatrix} \quad (\text{A.5.29})$$

and, similarly,

$$\text{var} \begin{pmatrix} s_*^2 \\ t_*^2 \end{pmatrix} = \begin{pmatrix} \frac{2\sigma^4}{N(n-1)} & 0 \\ 0 & \frac{2}{Nn} \left[\frac{\sigma^4}{n} + 2\sigma^2d + nd^2 \right] \end{pmatrix}. \quad (\text{A.5.30})$$

From its definition it follows that $s_*^2 \equiv \widehat{\sigma^2}$. The same is not true for t_*^2 . One reason to consider it nevertheless is its independence from s_*^2 . Indeed, $(\widehat{\mu}, s_*^2 \equiv \widehat{\sigma^2}, t_*^2)'$ is an estimator with mutually independent components. While the same is true when s^2 and t^2 are used instead,

the biases are larger.

For this case then, the choice between \hat{d} and t_*^2 is in terms of a trade-off between efficiency and independence.

To gauge the efficiency loss when using t_*^2 , the mean squared error is:

$$MSE(t_*^2) = \frac{2}{Nn} \left(\frac{\sigma^4}{n} + 2\sigma^2 d + nd^2 \right) + \frac{1}{n^2} \sigma^4,$$

and hence the relative MSE:

$$RMSE(t_*^2; \hat{d}) = \frac{2 \left(\frac{\sigma^4}{n} + 2\sigma^2 d + nd^2 \right) + \frac{N}{n} \sigma^4}{2 \left(\frac{\sigma^4}{n-1} + 2\sigma^2 d + nd^2 \right)}, \quad (\text{A.5.31})$$

which approaches infinity when N does, while n would remain constant. In other words, this estimator is inconsistent unless it is being applied in situations where n can also be considered to be large.

There are three distinct situations. First, when $N/n = \lambda n + o(n)$, for some λ , i.e., when N is of the order of n^2 , then, based on (A.5.31), the ARE is $2(d^2 + \lambda\sigma^4)/[2d^2]$. The magnitude of the efficiency loss depends on sizes of the parameters involved. Second, when $\mathcal{O}(N) < \mathcal{O}(n^2)$, the ARE equals 1. This includes the cases where N is constant, $N = n^{1/2}$, $N = n$, and $N = n^{3/2}$, for example. A constant or slowly increasing N is plausible in a meta-analytic context. Third, if N/n increases too quickly, i.e., $\mathcal{O}(N/n) > \mathcal{O}(n)$, then the estimator t_*^2 is inconsistent. This is the case, in particular, for bounded n .

The estimators s^2 and t^2 can be combined linearly to produce unbiased estimators. In other words, based on (A.5.23)–(A.5.24), the following corrections can be applied to (A.5.21)–(A.5.22):

$$s_{\text{corr}}^2 = \frac{n}{n-1}s^2, \quad (\text{A.5.32})$$

$$t_{\text{corr}}^2 = \frac{N}{N-1}t^2 - \frac{N}{(n-1)(N-1)}s^2. \quad (\text{A.5.33})$$

Interestingly, this requirement reproduces (A.5.27)–(A.5.28): the requirement of an unbiased estimator reproduces $\widehat{\sigma}^2$ and \widehat{d} , presented in (A.5.16)–(A.5.18) and hence also with their variance.

A.6 Details about the first simulation study

The simulation study, summarized in Section 7, is described in detail here.

A.6.1 Simulation method

The design of the simulation study is as follows.

- Each generated set of data consists of c_k clusters of size n_k , for $k = 1, \dots, K$. We choose $K = 4$ throughout.
- For a generated set of data, the splitting is done by placing all clusters of a given size in one sub-sample.
- The CS model parameters are $\mu = 0$, $d = 1$, and $\sigma^2 = 2$.

- After estimating the three model parameters within each sub-sample, they are combined using the following weighting methods: (a) equal, (b) proportional, where the weights are

$$w_k = \frac{c_k}{\sum_{\ell=1}^4 c_\ell},$$

and (c) size-proportional, where the weights for μ and d are:

$$w_k = \frac{c_k n_k}{\sum_{\ell=1}^4 c_\ell n_\ell},$$

while for σ^2 we take:

$$w_k = \frac{c_k(n_k - 1)}{\sum_{\ell=1}^4 c_\ell(n_\ell - 1)}.$$

- Per setting, 100 replications are considered.

These settings are applied to various combinations of the n_k and c_k , now described in turn.

A.6.2 Setting 1: equal $c_k \cdot n_k$, different c_k and n_k .

Consider 150 samples in each split, as follows: $(c_1, n_1) = (3, 50)$, $(c_2, n_2) = (5, 30)$, $(c_3, n_3) = (10, 15)$, and $(c_4, n_4) = (15, 10)$. The results are presented in Table A.6.1. Graphical depictions can be found in Figures A.6.1 and A.6.2. Figure A.6.1 shows that there is a different amount of information in the various sub-samples. This is not a problem, rather a consequence of the way the splits are created and the different

amounts of information carried in each. It reminds us that we need to be judicious how the information from the splits will be weighted. It is not a surprise that equal weights are a poor choice. The other methods perform similarly, and all do very well. To varying degrees, the same will be seen in Settings 2 and 3.

Table A.6.1: First simulation study. Setting 1. Average of split-specific and combined (weighted) parameters and their precision estimates.

	μ	$\text{var}(\mu)$	d	$\text{var}(d)$	σ^2	$\text{var}(\sigma^2)$
split1	-0.00396	0.05779	0.68143	0.41395	1.98676	0.00296
split2	0.05697	0.03071	0.80997	0.19712	1.98578	0.00304
split3	-0.02111	0.01174	0.95161	0.07869	1.97690	0.00319
split4	0.01123	0.00626	0.98870	0.04677	1.98056	0.00347
Equal	0.01078	0.03769	0.85793	0.09988	1.98250	0.01406
Prop	0.00698	0.03230	0.92245	0.08568	1.98081	0.01907
Size prop	0.01078	0.03769	0.85793	0.09988	1.98260	0.01405
Full	0.00780	0.03513	0.98016	0.08614	1.98257	0.01392

A.6.3 Setting 2: different $c_k \cdot n_k$, equal c_k , different n_k

To see the effect of split size, the following choices are made: $(c_1, n_1) = (4, 25)$, $(c_2, n_2) = (4, 50)$, $(c_3, n_3) = (4, 125)$, and $(c_4, n_4) = (4, 250)$. As a consequence, the size of the splits will be 100, 200, 500, and 1000, respectively. Table A.6.2 summarized the results, with graphical displays presented in Figures A.6.3 and A.6.4.

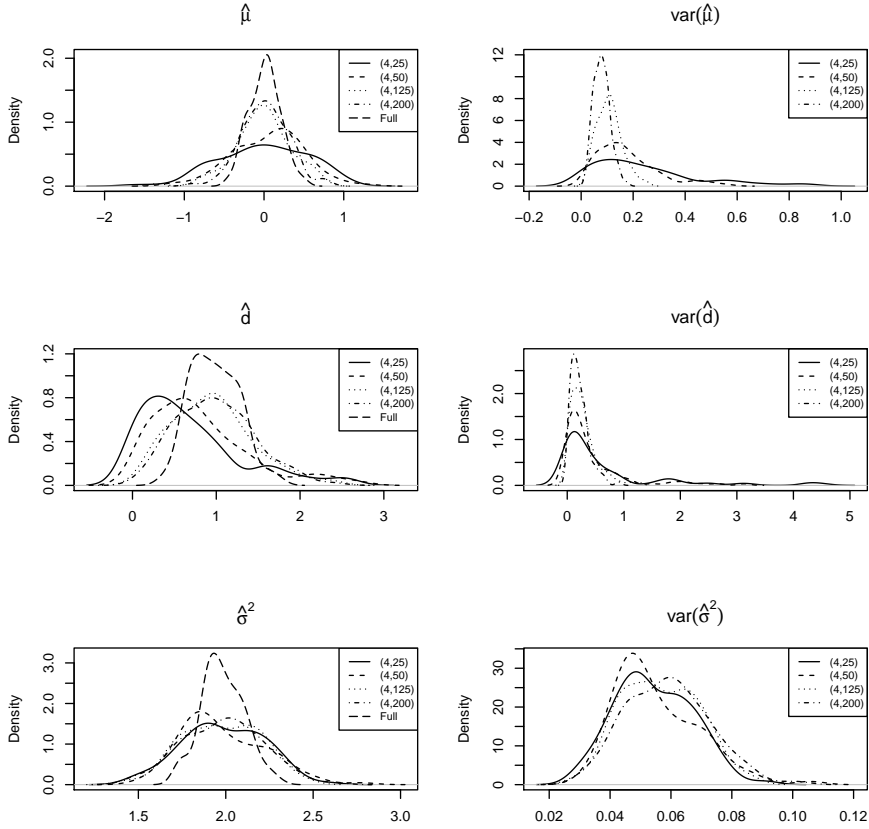


Figure A.6.1: First simulation study. Setting 1. Split-specific results.

A.6.4 Setting 3: different $c_k \cdot n_k$, different c_k , equal n_k

We now choose: $(c_1, n_1) = (10, 20)$, $(c_2, n_2) = (20, 20)$, $(c_3, n_3) = (50, 20)$, and $(c_4, n_4) = (100, 20)$. Table A.6.3 summarizes the results. Graphs can be found in Figures A.6.5 and A.6.6.

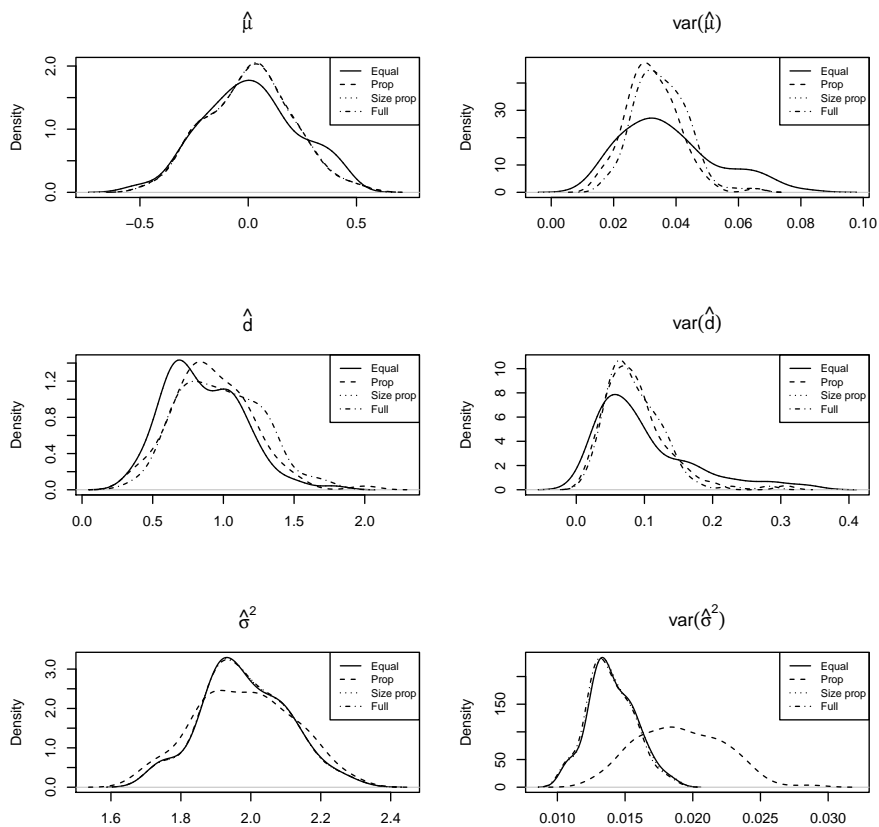


Figure A.6.2: First simulation study. Setting 1. Combining the results from the four splits, using equal, proportional, and size proportional weights. This is compared with full maximum likelihood.

A.6.5 Optimal, approximate optimal, and iterated optimal weights

Optimal weights were discussed in Section 6.1.1. When we plug the MLE's into the optimal weights, the result of using these weights is the

Table A.6.2: First simulation study. Setting 2. Average of split-specific and combined (weighted) parameters and their precision estimates.

	μ	$\text{var}(\mu)$	d	$\text{var}(d)$	σ^2	$\text{var}(\sigma^2)$
split1	-0.02515	0.05227	0.83440	0.52423	2.00347	0.00730
split2	0.01287	0.05157	0.86891	0.57904	1.97285	0.00160
split3	0.06812	0.03586	0.74147	0.23681	2.00165	0.00026
split4	-0.03676	0.02979	0.68241	0.14117	1.99216	0.00006
Equal	0.00477	0.05111	0.78180	0.14770	1.99253	0.00935
Prop	0.00477	0.05111	0.78180	0.14770	1.99253	0.00935
Size prop	-0.00147	0.07139	0.72798	0.16585	1.99328	0.00447
Full	0.00530	0.06339	0.89599	0.14604	1.99333	0.00446

MLE's itself. Of course, this is a circular reasoning, which is why one needs to resort to, for example, the approximate or iterated optimal weights derived in Section 6.1.2. For both of these, using Settings 1–3, we conducted simulations. They are reported in Figures A.6.7–A.6.9.

It is noteworthy that the behavior of the iterated optimal weights depends on c_k and n_k . First, they often but not always converge in a single iteration; the maximum number of iterations observed in our simulations being 6. Second, the iterated optimal weights converge to size optimal weights for σ^2 and to proportional weights for d .

Taken together, it follows that both approximately optimal and iterated optimal weights provide excellent results. The specific attraction of the approximate optimal weights is that they obviate the need for iteration, which is a factor of stability and speed.

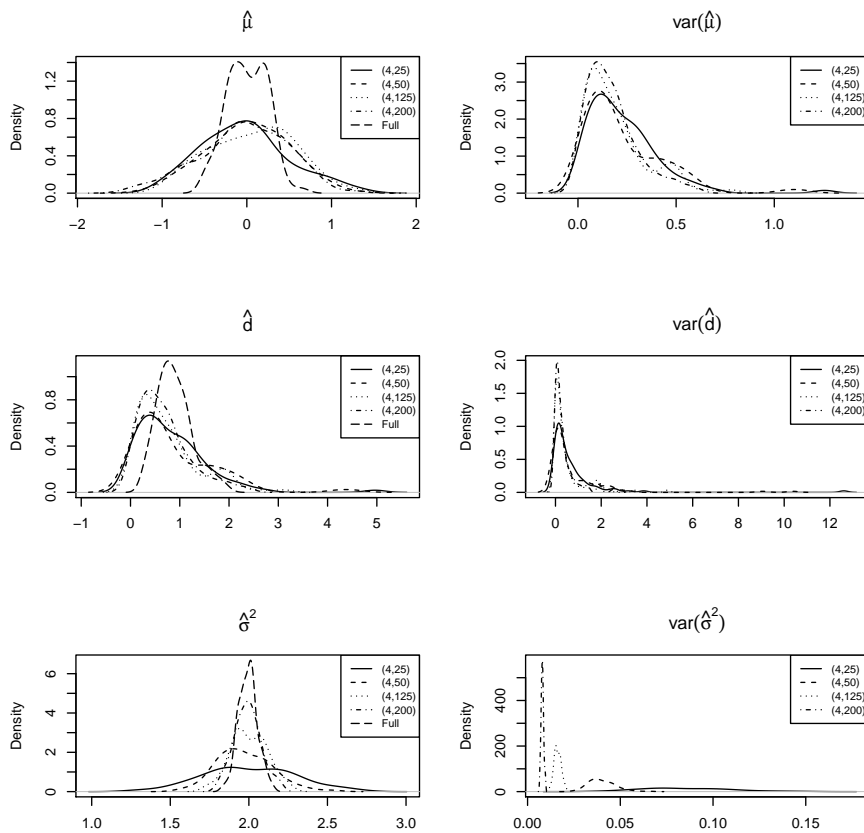


Figure A.6.3: First simulation study. Setting 2. Split-specific results.

A.7 Details about the second simulation study

The aim of this study is to compare the proposed method to two alternatives:

1. full maximum likelihood;

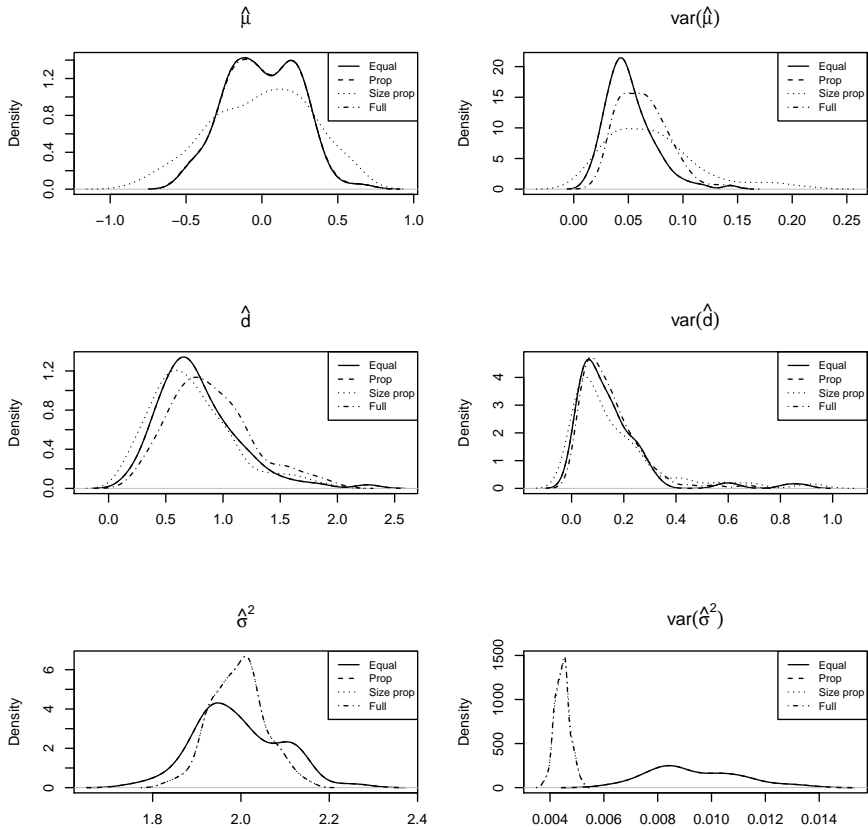


Figure A.6.4: First simulation study. Setting 2. Combining the results from the four splits, using equal, proportional, and size proportional weights. This is compared with full maximum likelihood.

2. the proposed sample-splitting method, allowing for closed forms;
3. using multiple imputation (MI) first, to render the clusters of equal sizes, and then apply closed-form solutions to the augmented

Table A.6.3: First simulation study. Setting 3. Average of split-specific and combined (weighted) parameters and their precision estimates.

	μ	$\text{var}(\mu)$	d	$\text{var}(d)$	σ^2	$\text{var}(\sigma^2)$
split1	0.00343	0.00900	0.84739	0.05169	2.02445	0.00190
split2	0.02553	0.00304	1.00224	0.01754	2.01962	0.00047
split3	-0.00010	0.00045	0.95794	0.00212	1.99765	0.00007
split4	0.01151	0.00012	1.01226	0.00064	1.98944	0.00002
Equal	0.01009	0.01139	0.95496	0.02694	2.00779	0.00486
Prop	0.00939	0.00604	0.98690	0.01369	1.99702	0.00234
Size prop	0.00939	0.00604	0.98690	0.01369	1.99702	0.00234
Full	0.00939	0.00614	1.00487	0.01372	1.99702	0.00233

balanced data, together with the combination rules.

A.7.1 Simulation plan

In order to study the effect of cluster sizes (n_k) and number of clusters of each size (c_k), 5 different configurations are considered:

Config. 1. $c_k = (15, 25, 30, 20, 10)$, $n_k = (8, 5, 3, 9, 15)$;

Config. 2. $c_k = (150, 250, 300, 200, 100)$, $n_k = (8, 5, 3, 9, 15)$;

Config. 3. $c_k = (1500, 2500, 3000, 2000, 1000)$, $n_k = (8, 5, 3, 9, 15)$;

Config. 4. $c_k = (15, 25, 30, 20, 10)$, $n_k = (80, 50, 30, 90, 150)$;

Config. 5. $c_k = (15, 25, 30, 20, 10)$, $n_k = (800, 500, 300, 900, 1500)$.

Each configuration is repeated 100 times.

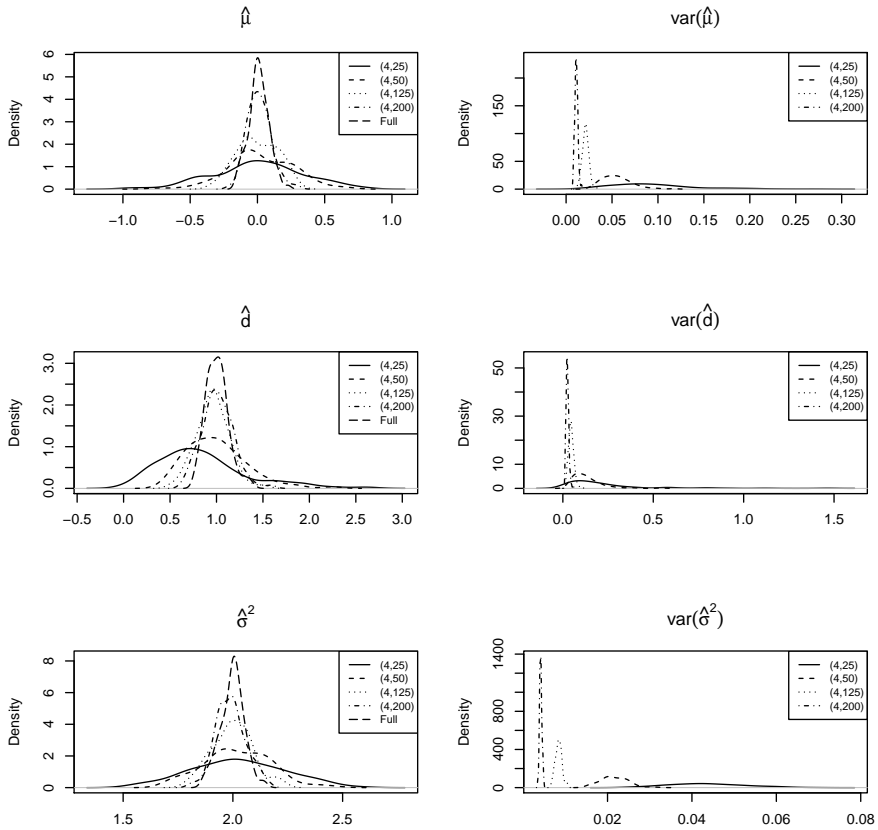


Figure A.6.5: First simulation study. Setting 3. Split-specific results.

Each cluster is generated from a CS model with $\mu_0 = 0$, $d_0 = 1$, and $\sigma_0^2 = 4$.

For estimating the parameters using the full unbalanced data, PROC MIXED in SAS (Version 9.4) is used with the covariance structure in the REPEATED statement set to `type=cs`.

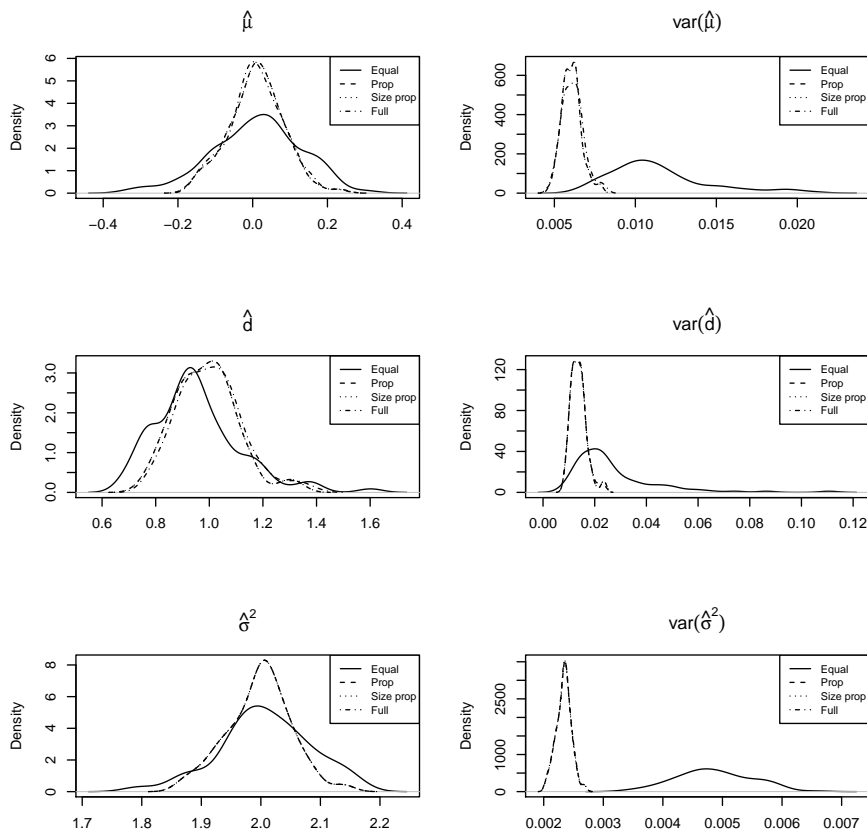


Figure A.6.6: First simulation study. Setting 3. Combining the results from the four splits, using equal, proportional, and size proportional weights. This is compared with full maximum likelihood.

The closed form solutions and their variances are implemented in R in three different ways. First, the formulas are implemented directly using ‘for’ loops. Following the ideas in Sikorska et al. (2013), it might be faster to replace ‘for’ loops with vectorized computation. For $\hat{\mu}_k$

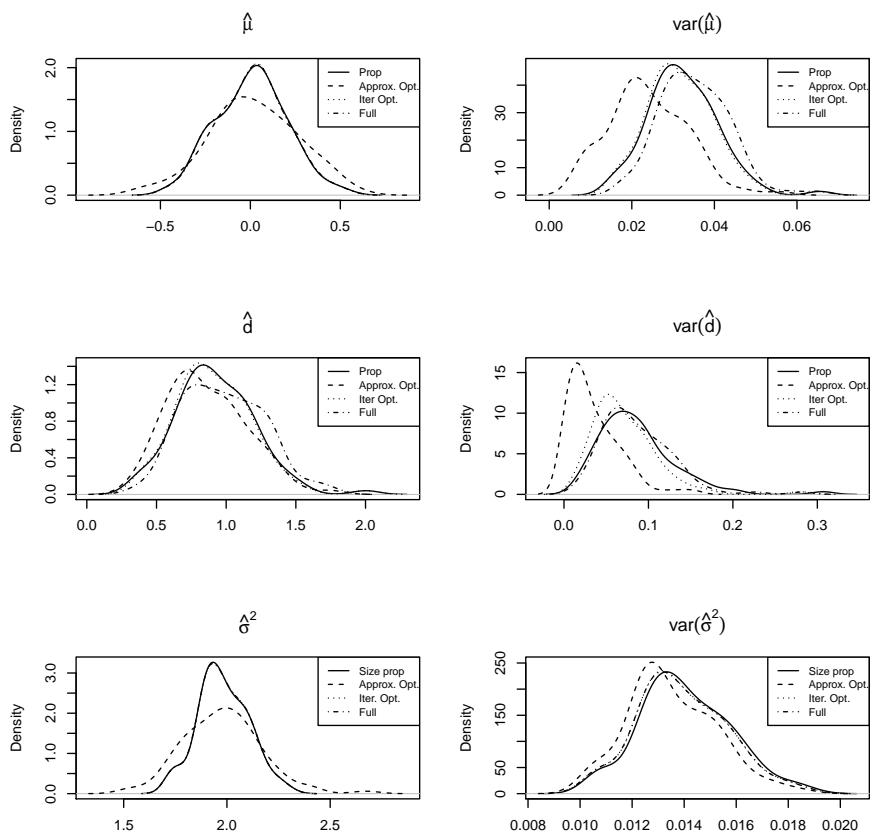


Figure A.6.7: First simulation study. Setting 1. (Size) proportional, approximate, and iterated optimal weights, as well as full maximum likelihood.

it is straightforward, since one just needs to compute an arithmetic average. If Z is a n_k times c_k matrix with its i th column defined as $Z_i^{(k)} = (Y_i^{(k)} - \mu_k \mathbf{1}_{n_k})$, then computing $\sum_{i=1}^{c_k} Z_i^{(k)'} Z_i^{(k)}$ is equivalent to replacing each element in matrix Z by its square, and then sum over the

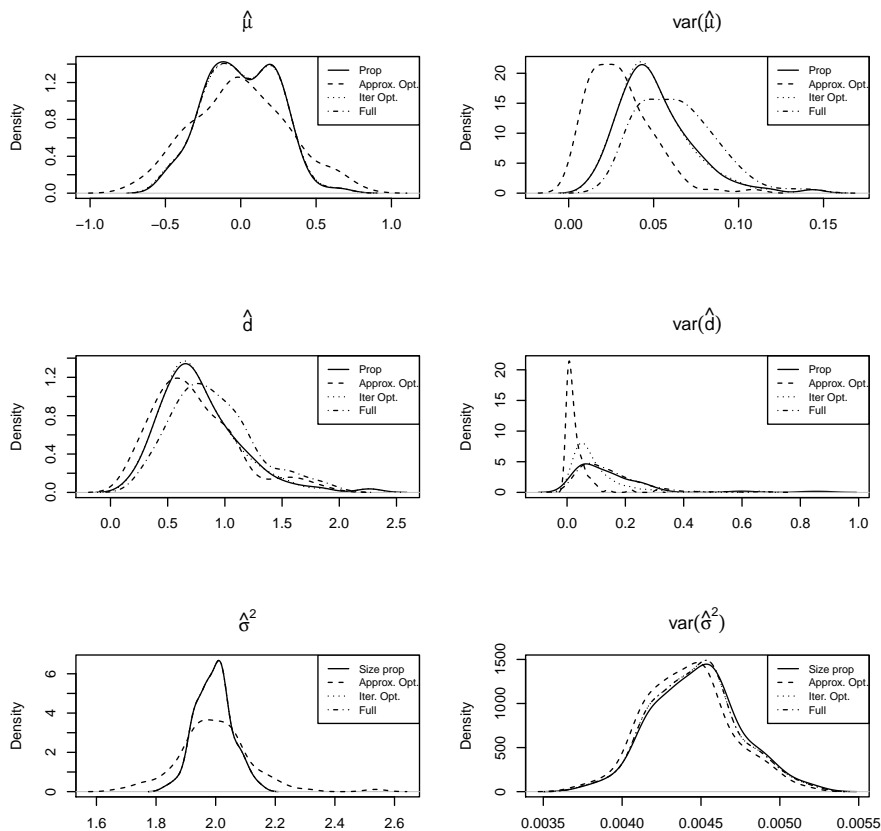


Figure A.6.8: First simulation study. Setting 2. (Size) proportional, approximate, and iterated optimal weights, as well as full maximum likelihood.

sum of its columns. Furthermore, $J_{n_k} Z_i^{(k)}$ would simply compute the sum of columns in matrix Z . Therefore, $\sum_{i=1}^{c_k} Z_i^{(k)'} J_{n_k} Z_I^{(k)}$ is equivalent to post-multiplying Z by the sum of its columns and then sum over this vector. In this way, within each split, the parameters can be estimated

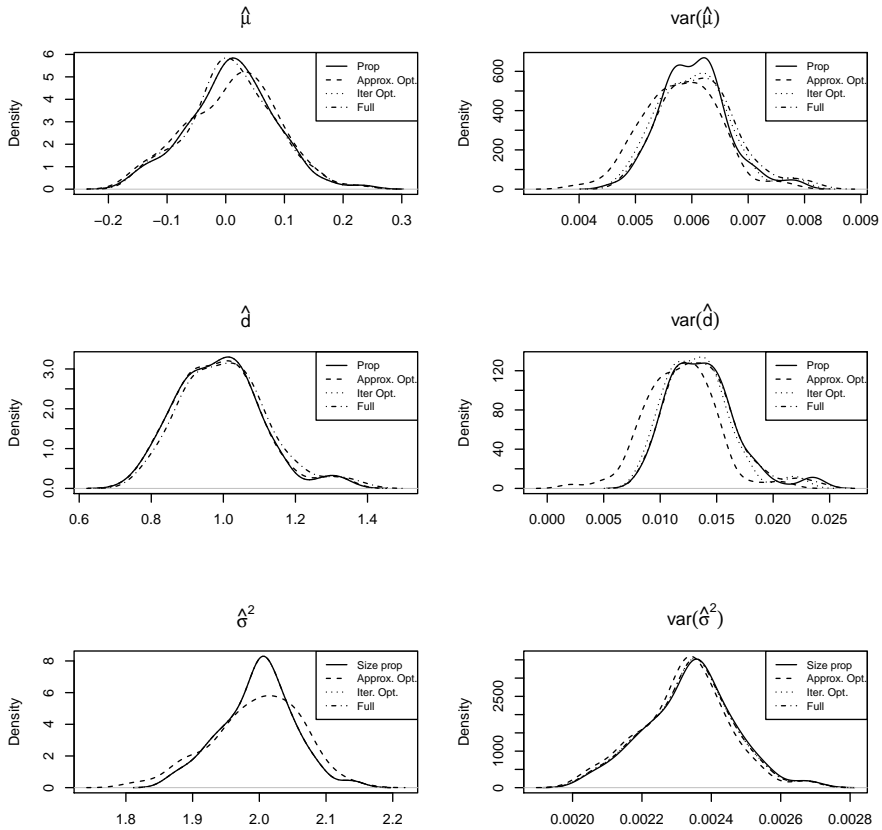


Figure A.6.9: First simulation study. Setting 3. (Size) proportional, approximate, and iterated optimal weights, as well as full maximum likelihood.

avoiding ‘for’ loops.

Second, it is also possible to find the estimates for all of the splits at once instead of computing them separately.

A third way consists of calculating all the estimates together and not split by split in a ‘for’ loop. This approach is possible via imposing balance through adding missing values in the matrix but, when multiplying and summing, ignoring the missing values. This is very easy in R.

We will compare computation time between these three approaches, for the five configurations.

Additionally, to combine the results from sample splitting, the same weights as used in the case study are considered here as well: equal, proportional, approximate scalar, scalar, and approximate optimal. In the case of the approximate optimal weights both simple and proper variances are calculated.

For the multiple imputation based approach, $M = 20$ imputations are considered and the conventional combination rules applied.

Note that the MI approach cannot be used with configurations 1, 4, and 5, because the number of available subjects in the observed dataset is less than the number of repeated measurements, leading to a singular covariance matrix. From the remaining configurations 2 and 3, we have chosen #2, which implies smaller numbers and hence is more challenging.

For each configuration we report three results: the estimated parameters, their standard errors, and the mean square error (MSE). Furthermore, we report computation time.

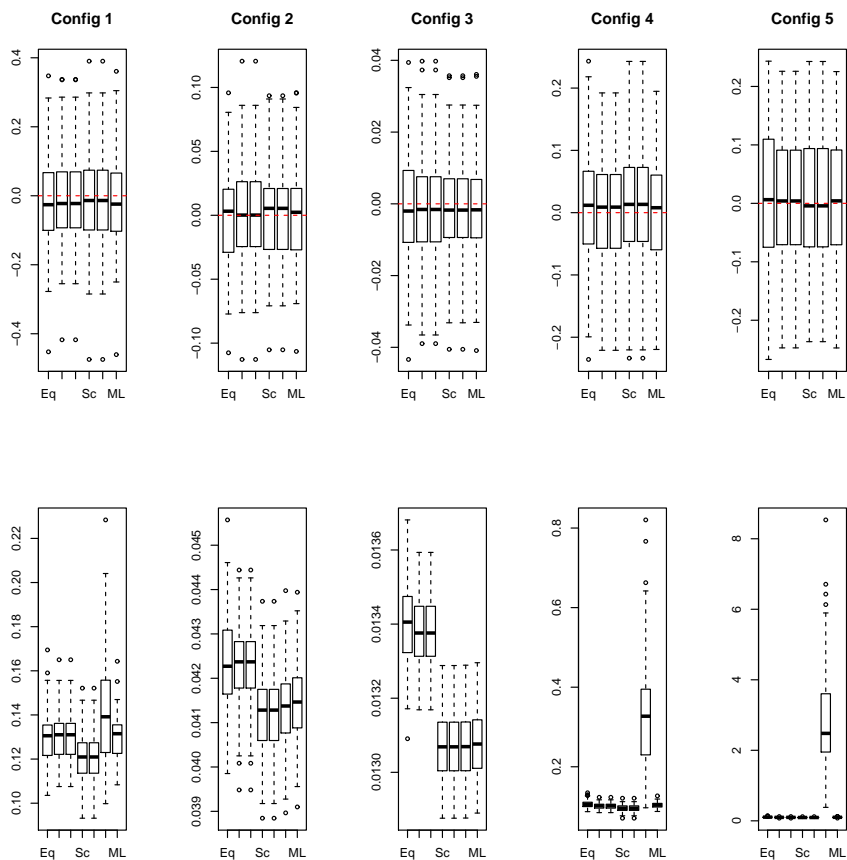


Figure A.7.10: Second simulation study. Estimates for μ (first row) and its standard error (second row).

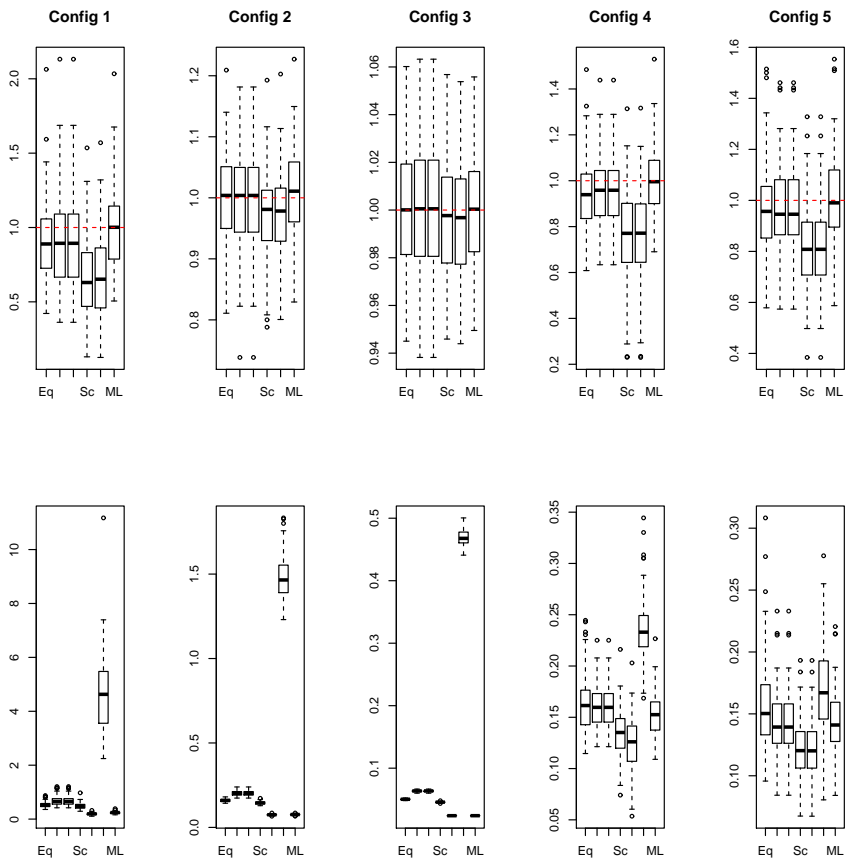


Figure A.7.11: Second simulation study. Estimates for d (first row) and standard errors (second row).

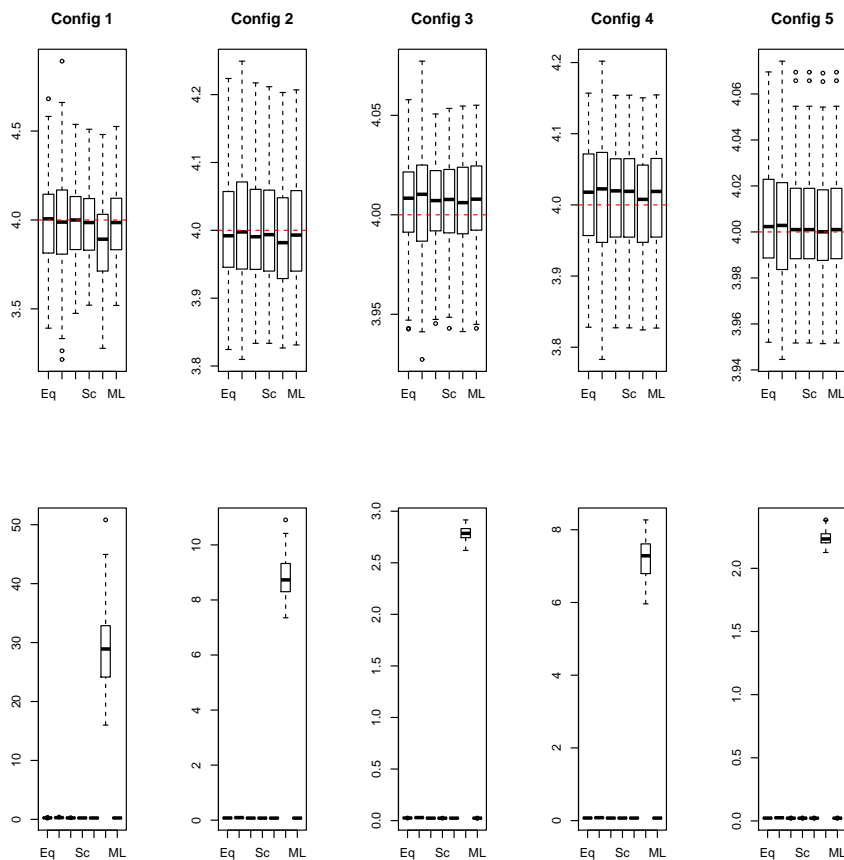


Figure A.7.12: Second simulation study. Estimates for σ^2 (first row) and standard errors (second row).

Table A.7.4: Second simulation study. Mean, standard deviation (S.D.) and MSE for μ among 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop	Approx. sc.	Scalar	Approx. opt.	ML
1	Mean	-1.72277E-02	-1.51605E-02	-1.51605E-02	-1.21296E-02	-1.21296E-02	-1.56028E-02
	S.D.	(1.32751E-01)	(1.28989E-01)	(1.28989E-01)	(1.32435E-01)	(1.32435E-01)	(1.29150E-01)
	MSE	1.77434E-02	1.67015E-02	1.67015E-02	1.75108E-02	1.75108E-02	1.67564E-02
2	Mean	-9.39926E-04	6.93419E-04	6.93419E-04	1.27613E-03	1.27613E-03	7.59533E-04
	S.D.	(3.93526E-02)	(3.95534E-02)	(3.95534E-02)	(3.84321E-02)	(3.84321E-02)	(3.83996E-02)
	MSE	1.53403E-03	1.54930E-03	1.54930E-03	1.46389E-03	1.46389E-03	1.46036E-03
3	Mean	-8.30934E-04	-1.25609E-03	-1.25609E-03	-1.26810E-03	-1.26810E-03	-1.31356E-03
	S.D.	(1.44839E-02)	(1.47545E-02)	(1.47545E-02)	(1.41310E-02)	(1.41310E-02)	(1.41704E-02)
	MSE	2.08376E-04	2.17095E-04	2.17095E-04	1.99298E-04	1.99298E-04	2.00517E-04
4	Mean	9.30928E-03	2.53713E-03	2.53713E-03	8.29367E-03	8.29367E-03	2.82086E-03
	S.D.	(9.26881E-02)	(8.32009E-02)	(8.32009E-02)	(9.76672E-02)	(9.76672E-02)	(8.28999E-02)
	MSE	8.59183E-03	6.85960E-03	6.85960E-03	9.51227E-03	9.51227E-03	6.81163E-03
5	Mean	9.77532E-03	1.02173E-02	1.02173E-02	8.72381E-03	8.72381E-03	1.02422E-02
	S.D.	(1.09769E-01)	(1.04876E-01)	(1.04876E-01)	(1.04982E-01)	(1.04982E-01)	(1.04847E-01)
	MSE	1.20243E-02	1.09934E-02	1.09934E-02	1.09872E-02	1.09872E-02	1.09880E-02

A.7.2 Simulation results

Based on the simulation results, it appears that using equal weights is not recommended, while using proportional weights produces results comparable with ML. Of course, in case of σ^2 the approximate scalar weights work better comparing with ML. An interesting outcome of the simulation is that by keeping the number of clusters of different sizes constant, but allowing the cluster sizes to increase, improves estimation of σ^2 , while increasing the number of clusters and keeping their sizes constant improves the estimation of d . This is not surprising, because d is the between-cluster variability, which is easier to estimate from a larger number of clusters. This should be seen against the background of relatively small differences anyway.

The results based on MI are not comparable with sample-splitting results.

Table A.7.5: Second simulation study. Mean and standard deviation (S.D.) for standard errors of μ estimates in 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop	Approx. sc.	Scalar	Simple opt.	Proper opt.	ML
1	Mean	1.29844E-01	1.29733E-01	1.29733E-01	1.20593E-01	1.20593E-01	1.41085E-01	1.29648E-01
	S.D.	(1.18175E-02)	(1.07496E-02)	(1.07496E-02)	(1.09122E-02)	(1.09122E-02)	(2.25155E-02)	(9.95614E-03)
2	Mean	4.23655E-02	4.23056E-02	4.23056E-02	4.11657E-02	4.11657E-02	4.12635E-02	4.14298E-02
	S.D.	(1.08386E-03)	(8.88747E-04)	(8.88747E-04)	(8.71465E-04)	(8.71465E-04)	(8.90259E-04)	(8.66432E-04)
3	Mean	1.34008E-02	1.33822E-02	1.33822E-02	1.30725E-02	1.30725E-02	1.30728E-02	1.30799E-02
	S.D.	(1.21880E-04)	(9.92324E-05)	(9.92324E-05)	(9.68871E-05)	(9.68871E-05)	(9.68974E-05)	(9.62652E-05)
4	Mean	1.06373E-01	1.01232E-01	1.01232E-01	9.60807E-02	9.60807E-02	3.30358E-01	1.03382E-01
	S.D.	(9.80278E-03)	(7.26031E-03)	(7.26031E-03)	(8.36242E-03)	(8.36242E-03)	(1.39042E-01)	(7.17708E-03)
4	Mean	1.05176E-01	9.84615E-02	9.84615E-02	9.45066E-02	9.45066E-02	2.81427E+00	1.00533E-01
	S.D.	(1.10711E-02)	(8.18259E-03)	(8.18259E-03)	(8.14229E-03)	(8.14229E-03)	(1.41211E+00)	(8.28537E-03)

In particular, the variance component d is underestimated using MI, while σ^2 is overestimated. The larger standard errors in this case suggest that the sample-splitting methods use information more efficiently.

Comparing computation times, the closed-form approaches are the clear winners, further enhanced by smaller standard errors. Furthermore, it follows that computing the estimates in a semi-parallel fashion, thus avoiding ‘for’ loops within the splits but using then between splits, is most efficient. Unless the n_k ’s are very large, computing all estimates at once is the most efficient. Of course, if the estimates for different splits can be done in parallel (without ‘for’ loops), this is more efficient than estimating them all at once.

A.8 The balanced conditionally independent model

In this case, one imposes the following structure on (??):

Table A.7.6: Second simulation study. Mean, standard deviation (S.D.) and MSE for d estimates in 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop	Approx. sc.	Scalar	Approx. opt	ML
1	Mean	9.09580E-01	9.00788E-01	9.00788E-01	6.63293E-01	6.71111E-01	9.86549E-01
	S.D.	(2.66885E-01)	(2.99736E-01)	(2.99736E-01)	(2.70556E-01)	(2.79060E-01)	(2.55874E-01)
	MSE	7.86910E-02	9.87862E-02	9.87862E-02	1.85840E-01	1.85264E-01	6.49975E-02
2	Mean	9.98984E-01	9.97534E-01	9.97534E-01	9.72269E-01	9.73771E-01	1.00712E+00
	S.D.	(7.44958E-02)	(8.28143E-02)	(8.28143E-02)	(7.28186E-02)	(7.22762E-02)	(7.08463E-02)
	MSE	5.49515E-03	6.79570E-03	6.79570E-03	6.01854E-03	5.85958E-03	5.01969E-03
3	Mean	1.00004E+00	9.99697E-01	9.99697E-01	9.97218E-01	9.97153E-01	1.00019E+00
	S.D.	(2.61046E-02)	(2.82480E-02)	(2.82480E-02)	(2.48989E-02)	(2.48068E-02)	(2.44715E-02)
	MSE	6.74637E-04	7.90063E-04	7.90063E-04	6.21496E-04	6.17332E-04	5.92900E-04
4	Mean	9.40257E-01	9.50756E-01	9.50756E-01	7.65219E-01	7.65362E-01	9.96005E-01
	S.D.	(1.53414E-01)	(1.48216E-01)	(1.48216E-01)	(1.91865E-01)	(1.91614E-01)	(1.49451E-01)
	MSE	2.68697E-02	2.41734E-02	2.41734E-02	9.15661E-02	9.14038E-02	2.21282E-02
5	Mean	9.63459E-01	9.68182E-01	9.68182E-01	8.21266E-01	8.21270E-01	1.00958E+00
	S.D.	(1.73767E-01)	(1.63471E-01)	(1.63471E-01)	(1.72162E-01)	(1.72163E-01)	(1.69235E-01)
	MSE	3.12283E-02	2.74677E-02	2.74677E-02	6.12894E-02	6.12881E-02	2.84459E-02

- $X_i^{(k)}$ can be rewritten in terms of a first matrix that imposes structure between clusters (e.g., treatment effect), termed $A_i^{(k)}$, and a second one that imposes structure within clusters (e.g., time evolution), $T_i^{(k)'} = (Z_i^{(k)'}, Q_i^{(k)'})'$.
- The matrices $A_i^{(k)}$, $Z_i^{(k)}$, and $Q_i^{(k)}$ are constant among all clusters of size n_k .
- The matrix $\Sigma_i^{(k)} = \sigma^2 I_{n_k}$.

This is the general, balanced growth-curve model as studied by Lange and Laird (1989) and Verbeke and Fieuws (2007). Building on their development, we will now derive sufficient statistics and associated maximum likelihood estimators for the parameters in this model. This

Table A.7.7: Second simulation study. Mean and standard deviation (S.D.) for standard errors of d estimates in 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop.	Approx. sc.	Scalar	Simple opt.	Proper opt.	ML
1	Mean	5.30888E-01	6.76159E-01	6.76159E-01	4.83519E-01	1.95328E-01	4.66561E+00	2.36408E-01
	S.D.	(1.00495E-01)	(1.66365E-01)	(1.66365E-01)	(1.10589E-01)	(4.01066E-02)	(1.40173E+00)	(3.91752E-02)
2	Mean	1.60365E-01	2.03126E-01	2.03126E-01	1.45284E-01	7.48056E-02	1.47955E+00	7.60798E-02
	S.D.	(8.29444E-03)	(1.42616E-02)	(1.42616E-02)	(8.70034E-03)	(3.35411E-03)	(1.31587E-01)	(3.26483E-03)
3	Mean	5.02596E-02	6.34861E-02	6.34861E-02	4.56061E-02	2.39463E-02	4.67989E-01	2.39748E-02
	S.D.	(8.43976E-04)	(1.44201E-03)	(1.44201E-03)	(8.34510E-04)	(3.68251E-04)	(1.24414E-02)	(3.66865E-04)
4	Mean	1.62410E-01	1.59656E-01	1.59656E-01	1.34256E-01	1.24552E-01	2.35351E-01	1.51740E-01
	S.D.	(2.83009E-02)	(2.05482E-02)	(2.05482E-02)	(2.31559E-02)	(2.51091E-02)	(3.00109E-02)	(2.11541E-02)
5	Mean	1.54122E-01	1.43313E-01	1.43313E-01	1.22117E-01	1.22024E-01	1.70868E-01	1.43764E-01
	S.D.	(3.48725E-02)	(2.50371E-02)	(2.50371E-02)	(2.30972E-02)	(2.31064E-02)	(3.70951E-02)	(2.38992E-02)

can be expressed

$$\mathbf{Y} = A(\beta_1, \beta_2) \begin{pmatrix} Z \\ Q \end{pmatrix} + BZ + \epsilon.$$

Here, \mathbf{Y} is an $N \times n$ matrix stacking the outcomes of all clusters of size c , A , Z , and Q group the designs mentioned in Section ??, the vectors β_1 and β_2 contain the fixed effects, B contains N rows of length q , representing the q -dimensional random-effects vector, and ϵ shares its dimensions with \mathbf{Y} .

Now, define K the projection matrix such that $K'K = I_{r-q}$, for an appropriate dimension r , and $ZK = 0$. Then, set $P = QK$ and consider the projection model:

$$\mathbf{Y}_1 \equiv YK = A\beta_2P + \epsilon K.$$

The variance of a cluster is $\sigma^2 I_{r-q}$. Next, define H such that $H'H = I_q$

Table A.7.8: Second simulation study. Mean, standard deviation (S.D.) and MSE for σ^2 estimates in 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop.	Approx. sc.	Scalar	Approx. opt.	ML
1	Mean	3.98608E+00	3.99739E+00	3.98571E+00	3.98364E+00	3.87487E+00	3.98075E+00
	S.D.	(2.50882E-01)	(2.95821E-01)	(2.35138E-01)	(2.33650E-01)	(2.38167E-01)	(2.30477E-01)
	MSE	6.25062E-02	8.66420E-02	5.49414E-02	5.43140E-02	7.18141E-02	5.29588E-02
2	Mean	4.00184E+00	4.00681E+00	4.00177E+00	4.00087E+00	3.99027E+00	4.00064E+00
	S.D.	(7.85190E-02)	(9.05849E-02)	(7.57443E-02)	(7.56486E-02)	(7.50477E-02)	(7.51739E-02)
	MSE	6.10698E-03	8.16998E-03	5.68296E-03	5.66624E-03	5.67055E-03	5.59502E-03
3	Mean	4.00509E+00	4.00491E+00	4.00575E+00	4.00590E+00	4.00472E+00	4.00587E+00
	S.D.	(2.45760E-02)	(2.82395E-02)	(2.36027E-02)	(2.36983E-02)	(2.37694E-02)	(2.36697E-02)
	MSE	6.23828E-04	8.13646E-04	5.84605E-04	5.90806E-04	5.81652E-04	5.89081E-04
4	Mean	4.01402E+00	4.01190E+00	4.01302E+00	4.01304E+00	4.00363E+00	4.01306E+00
	S.D.	(7.69655E-02)	(8.78962E-02)	(7.32088E-02)	(7.31202E-02)	(7.20589E-02)	(7.31207E-02)
	MSE	6.06101E-03	7.79021E-03	5.47558E-03	5.46319E-03	5.15372E-03	5.46370E-03
5	Mean	4.00346E+00	4.00338E+00	4.00292E+00	4.00292E+00	4.00192E+00	4.00292E+00
	S.D.	(2.56561E-02)	(2.79741E-02)	(2.46670E-02)	(2.46664E-02)	(2.46901E-02)	(2.46669E-02)
	MSE	6.63599E-04	7.86128E-04	6.10884E-04	6.10854E-04	6.07187E-04	6.10909E-04

and $QH = 0$. A second projection model emerges:

$$\mathbf{Y}_2 \equiv YH = A\beta_1 + B + \epsilon H.$$

The variance of a cluster is $\sigma^2 I_q + H'DH$, with D the variance-covariance matrix of the vector of random effects. Importantly, projections $\mathbf{Y}_1 \perp \mathbf{Y}_2$.

Conventional algebra leads from these to the following set of sufficient statistics:

$$T_1 = (A'A)^{-1}A'\mathbf{Y}_1P'(PP')^{-1}, \quad (\text{A.8.1})$$

$$T_2 = \text{tr} \left\{ \mathbf{Y}'_1 \left[I - A(A'A)^{-1}A' \right] \mathbf{Y}_1 \right\}, \quad (\text{A.8.2})$$

$$T_3 = (A'A)^{-1}A'\mathbf{Y}_2, \quad (\text{A.8.3})$$

$$T_4 = \mathbf{Y}'_2 \left[I - A(A'A)^{-1}A' \right] \mathbf{Y}_2. \quad (\text{A.8.4})$$

Table A.7.9: Second simulation study. Mean and standard deviation (S.D.) for standard errors of σ^2 estimates in 100 replications for each configuration using different combination weights comparing with full sample MLE.

Config.		Equal	Prop.	Approx. sc.	Scalar	Simple opt.	Proper opt.	ML
1	Mean	2.53758E-01	2.95524E-01	2.40482E-01	2.38699E-01	2.30855E-01	2.88730E+01	2.35804E-01
	S.D.	(1.98892E-02)	(3.33124E-02)	(1.45355E-02)	(1.39849E-02)	(1.38813E-02)	(6.93951E+00)	(1.34905E-02)
2	Mean	7.98210E-02	9.26570E-02	7.57927E-02	7.53242E-02	7.48395E-02	8.82494E+00	7.49872E-02
	S.D.	(1.84469E-03)	(3.08841E-03)	(1.45443E-03)	(1.42354E-03)	(1.40575E-03)	(6.92140E-01)	(1.39989E-03)
3	Mean	2.52202E-02	2.92034E-02	2.39687E-02	2.38353E-02	2.37400E-02	2.78559E+00	2.37444E-02
	S.D.	(1.85701E-04)	(3.12856E-04)	(1.42587E-04)	(1.40888E-04)	(1.40071E-04)	(6.70930E-02)	(1.39784E-04)
4	Mean	7.22378E-02	8.07793E-02	7.01684E-02	7.01663E-02	6.99970E-02	7.23107E+00	7.01238E-02
	S.D.	(1.54384E-03)	(2.35043E-03)	(1.28712E-03)	(1.28385E-03)	(1.26263E-03)	(5.28710E-01)	(1.27765E-03)
5	Mean	2.25782E-02	2.52054E-02	2.19702E-02	2.19702E-02	2.19647E-02	2.23651E+00	2.19689E-02
	S.D.	(1.55999E-04)	(2.29649E-04)	(1.35364E-04)	(1.35358E-04)	(1.35462E-04)	(5.50362E-02)	(1.35379E-04)

Sufficient statistics (A.8.1)–(A.8.4) lead to the maximum likelihood estimators:

$$\hat{\beta}_1 = T_1, \quad (\text{A.8.5})$$

$$\hat{\beta}_2 = T_3, \quad (\text{A.8.6})$$

$$\hat{\sigma}^2 = \frac{1}{N(n-q)} T_2, \quad (\text{A.8.7})$$

$$\hat{D} = \frac{1}{N} T_4 - \hat{\sigma}^2 I_q. \quad (\text{A.8.8})$$

Note that the estimators for the fixed effects do not involve the variance components.

A.9 Algebraic derivations in the AR(1) case

Here, we present more detailed derivations of the key algebraic expressions presented in Sections ?? and ??.

Table A.7.10: Second simulation study. Computation time (in seconds) using closed-form solutions with different implementation forms, compared to PROC MIXED.

Config.		Split by split		Together	PROC MIXED
		without loops	using loops		
1	Mean	0.00520	0.00980	0.00640	0.34155
	S.D.	(0.00882)	(0.00943)	(0.00823)	(0.17711)
2	Mean	0.02150	0.07340	0.05660	0.35575
	S.D.	(0.01617)	(0.02006)	(0.09662)	(0.03073)
3	Mean	0.17980	0.73480	0.43400	0.81783
	S.D.	(0.02292)	(0.05835)	(0.11543)	(0.02582)
4	Mean	0.00220	0.00610	0.00360	2.58808
	S.D.	(0.00579)	(0.01497)	(0.00689)	(0.40720)
5	Mean	0.04030	0.27490	0.02130	629.58333
	S.D.	(0.01521)	(0.01941)	(0.00872)	(116.09435)

A.9.1 Some useful expressions

Consider,

$$C = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots \\ & 1 & \rho & \rho^2 & \dots \\ & & \ddots & \ddots & \ddots \\ & & & 1 & \rho \\ & & & & 1 \end{pmatrix}, \quad (\text{A.9.1})$$

then,

$$\Sigma = \sigma^2 C. \quad (\text{A.9.2})$$

Table A.7.11: Second simulation study. Mean, standard deviation (S.D.) and MSE for CS parameter estimates in 100 replications for configuration 2 using different combination weights comparing with full sample MLE and MI-MLE.

		Equal	Prop.	Approx. sc.	Scalar	Approx. opt.	MI	ML
μ	Mean	-5.09643E-03	-2.62061E-03	-2.62061E-03	-2.88933E-03	-2.88933E-03	-8.04195E-03	-3.28224E-03
	S.D.	(4.85424E-02)	(4.91772E-02)	(4.91772E-02)	(4.68032E-02)	(4.68032E-02)	(6.00662E-02)	(4.71723E-02)
	MSE	2.35877E-03	2.40108E-03	2.40108E-03	2.17698E-03	2.17698E-03	3.63654E-03	2.21375E-03
d	Mean	9.98392E-01	9.95216E-01	9.95216E-01	9.70589E-01	9.71627E-01	3.51123E-02	9.92960E-01
	S.D.	(7.33193E-02)	(7.46946E-02)	(7.46946E-02)	(7.59516E-02)	(7.53362E-02)	(1.83983E-02)	(7.50610E-02)
	MSE	5.32455E-03	5.54638E-03	5.54638E-03	6.57598E-03	6.42380E-03	9.31343E-01	5.62737E-03
σ^2	Mean	4.00782E+00	4.00791E+00	4.00581E+00	4.00544E+00	3.99400E+00	5.32627E+00	4.00347E+00
	S.D.	(7.66500E-02)	(8.98881E-02)	(7.19708E-02)	(7.13441E-02)	(7.19080E-02)	(1.55770E-01)	(7.56968E-02)
	MSE	5.87763E-03	8.06169E-03	5.16181E-03	5.06871E-03	5.15505E-03	1.78301E+00	5.68472E-03

Table A.7.12: Second simulation study. Mean and standard deviation (S.D.) for the standard error of CS parameter estimates in 100 replications for configuration 2 using different combination weights comparing with full sample MLE and MI-MLE.

		Equal	Prop.	Approx. sc.	Scalar	Simple opt.	Proper opt.	MI	ML
μ	Mean	4.24162E-02	4.22766E-02	4.22766E-02	4.11356E-02	4.11356E-02	4.12439E-02	4.95431E-02	4.14004E-02
	S.D.	(1.21548E-03)	(8.59430E-04)	(8.59430E-04)	(9.31222E-04)	(9.31222E-04)	(9.42598E-04)	(7.65032E-03)	(9.11785E-04)
d	Mean	1.60545E-01	2.02922E-01	2.02922E-01	1.45623E-01	7.47531E-02	1.48865E+00	9.10211E-02	7.54391E-02
	S.D.	(8.88524E-03)	(1.47416E-02)	(1.47416E-02)	(9.84170E-03)	(3.76197E-03)	(1.29914E-01)	(5.18621E-03)	(3.38044E-03)
σ^2	Mean	7.99442E-02	9.26315E-02	7.58626E-02	7.54139E-02	7.49171E-02	8.86450E+00	1.64310E-01	7.50377E-02
	S.D.	(1.87358E-03)	(3.14842E-03)	(1.39531E-03)	(1.34244E-03)	(1.33807E-03)	(6.56714E-01)	(2.27104E-02)	(1.42688E-03)

It can be shown that:

$$\det(C) = \left(1 - \rho^2\right)^{n-1} \quad (\text{A.9.3})$$

The inverse of C can be calculated as follows:

$$C^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & \dots & 0 \\ -\rho & 1 + \rho^2 & \ddots & \dots \\ & \ddots & \ddots & \ddots \\ & & 1 + \rho^2 & -\rho \\ & & -\rho & 1 \end{pmatrix}, \quad (\text{A.9.4})$$

as one may see C^{-1} is a symmetric-tridiagonal matrix with constant diagonal except for the outer entries, and constant first off-diagonal.

Consider:

$$C^{-1} = \frac{1}{1 - \rho^2} G. \quad (\text{A.9.5})$$

Then, by taking the derivative with respect to ρ :

$$\frac{\partial C^{-1}}{\partial \rho} = \frac{2\rho}{(1 - \rho^2)^2} G + \frac{1}{1 - \rho^2} H, \quad (\text{A.9.6})$$

where, $H = \frac{\partial G}{\partial \rho}$ and has the form:

$$H = \begin{pmatrix} 0 & -1 & \dots & 0 \\ -1 & 2\rho & \ddots & \dots \\ \ddots & \ddots & \ddots & \ddots \\ & \ddots & 2\rho & -1 \\ 0 & & -1 & 0 \end{pmatrix}. \quad (\text{A.9.7})$$

Also, considering the fact $CC^{-1} = I$, one can derive:

$$\frac{\partial C^{-1}}{\partial \rho} = -C^{-1} \frac{\partial C}{\partial \rho} C^{-1}. \quad (\text{A.9.8})$$

A.9.2 The likelihood estimators in a given cluster

The likelihood function for an n_k -dimensional multivariate normal sample of size c_k has the following form:

$$L = \prod_{i=1}^{C_k} \frac{1}{|\Sigma|^{1/2} (2\pi)^{n_k/2}} \exp \left\{ -\frac{1}{2} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i) \right\}. \quad (\text{A.9.9})$$

Therefore, the non-constant terms of the log-likelihood are as follows:

$$\ell \propto -\frac{C_k}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^{C_k} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i), \quad (\text{A.9.10})$$

which considering (A.9.3) for AR(1):

$$|\Sigma| = (\sigma^2)^{n_k} (1 - \rho^2)^{n_k - 1}. \quad (\text{A.9.11})$$

As a general case, if we consider the mean as linear model with the form $\mu_i = X_i \beta$, one can derive:

$$\frac{\partial \ell}{\partial \mu_i} = \Sigma^{-1} \sum_{i=1}^{C_k} (y_i - \mu_i) = 0 \Rightarrow \hat{\beta} = (X'X)^{-1} X'y. \quad (\text{A.9.12})$$

Now expanding the log-likelihood for σ^2 and ρ , we have:

$$\ell \propto -\frac{C_k}{2} n_k \ln \sigma^2 - \frac{C_k}{2} (n_k - 1) \ln(1 - \rho^2) - \frac{1}{2} \sum_{i=1}^{C_k} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i). \quad (\text{A.9.13})$$

Considering $\Sigma = \sigma^2 C$ and (A.9.4), the derivative with respect to σ^2 is as follows:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{C_k n_k}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} \sum_{i=1}^{C_k} (y_i - \mu_i)' C^{-1} (y_i - \mu_i). \quad (\text{A.9.14})$$

Solving $\frac{\partial \ell}{\partial \sigma^2} = 0$ gives:

$$\hat{\sigma}^2 = \frac{1}{C_k n_k} \sum_{i=1}^{C_k} (y_i - \mu_i)' C^{-1} (y_i - \mu_i). \quad (\text{A.9.15})$$

One may notice that C^{-1} contains the parameter ρ .

Taking the derivative of (A.9.13) with respect to ρ gives:

$$\frac{\partial \ell}{\partial \rho} = \frac{C_k(n_k - 1)}{2} \frac{2\rho}{1 - \rho^2} - \frac{1}{\sigma^2} \sum_{i=1}^{C_k} (y_i - \mu_i)' \frac{\partial C^{-1}}{\partial \rho} (y_i - \mu_i). \quad (\text{A.9.16})$$

Setting $\frac{\partial \ell}{\partial \rho} = 0$ gives:

$$\hat{\sigma}^2 \frac{2\hat{\rho}}{1 - \hat{\rho}^2} = \frac{1}{C_k(n_k - 1)} \sum_{i=1}^{C_k} (y_i - \mu_i)' \frac{\partial C^{-1}}{\partial \rho} (y_i - \mu_i). \quad (\text{A.9.17})$$

Solving (A.9.15) and (A.9.17) gives $\hat{\sigma}^2$ and $\hat{\rho}$. For any $(n_k \times n_k)$ matrix Q , $\sum_i (y_i - \mu_i)' Q (y_i - \mu_i)$ equals $\text{tr} \{SQ\}$, where tr denotes the trace of a matrix, and $S = \sum_i (y_i - \mu_i)(y_i - \mu_i)'$. Hence, from (A.9.15), (A.9.17),

(A.9.5), (A.9.6), and (A.9.8), one can write:

$$\begin{cases} (1 - \hat{\rho}^2)\hat{\sigma}^2 = \frac{1}{C_k n_k} \text{tr}\{SG\}, \\ (1 - \hat{\rho}^2)\hat{\sigma}^2 = \frac{1}{C_k n_k} \text{tr}\{SG\} + \frac{1 - \hat{\rho}^2}{2\hat{\rho}} \frac{1}{C_k(n_k - 1)} \frac{1}{C_k(n_k - 1)} \text{tr}\{SH\}. \end{cases} \quad (\text{A.9.18})$$

Set $g = \text{tr}\{SG\}$ and $h = \text{tr}\{SH\}$, it follows that

$$\frac{g}{n_k} + \frac{1 - \hat{\rho}^2}{2\hat{\rho}} h = 0. \quad (\text{A.9.19})$$

Given that both g and h are functions of ρ only, ρ can be estimated using (A.9.19). Given ρ , one can use one of equations in (A.9.18) to estimate σ^2 .

Let us consider some special cases. For $n_k = 2$:

$$G = \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

Therefore, g and h can be computed as:

$$g = \text{tr} \left[\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \right] = S_{11} - 2\rho S_{12} + S_{22}.$$

$$h = \text{tr} \left[\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \right] = -2S_{12}.$$

Now using (A.9.19):

$$\hat{\rho}(S_{11} - 2\hat{\rho}S_{12} + S_{22} + (1 - \hat{\rho}^2)(-2S_{12})),$$

which gives:

$$\hat{\rho} = \frac{2S_{12}}{S_{11} + S_{22}}. \quad (\text{A.9.20})$$

Then, using first equation in (A.9.18):

$$(1 - \hat{\rho}^2)\hat{\sigma}^2 = \frac{1}{2C_k}(S_{11} - 2\hat{\rho}S_{12} + S_{22}),$$

which gives:

$$\hat{\sigma}^2 = \frac{S_{11} + S_{22}}{2C_k}. \quad (\text{A.9.21})$$

For $n_k = 3$:

$$\begin{aligned} g &= \text{tr} \left[\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1 + \rho^2 - \rho & \\ 0 & -\rho & 1 \end{pmatrix} \right] \\ &= S_{11} + S_{22} + S_{33} - 2\rho(S_{12} + S_{23}) + \rho^2 S_{22}, \\ h &= \text{tr} \left[\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ -1 & 2\rho & -1 \\ 0 & -1 & 0 \end{pmatrix} \right] \\ &= -2(S_{12} + S_{23}) + 2\rho S_{22}. \end{aligned}$$

Let,

$$\begin{cases} S = S_{11} + S_{22} + S_{33} \\ R = S_{12} + S_{23} \end{cases} \Rightarrow \begin{cases} g = S + \rho^2 S_{22} - 2\rho R \\ h = -2R + 2\rho S_{22} \end{cases}$$

Using (A.9.19):

$$2S_{22}\rho^3 - R\rho^2 - (S + 3S_{22})\rho + 3R = 0. \quad (\text{A.9.22})$$

Considering the results for $n_k = 2$ and $n_k = 3$, one can calculate (A.9.19) for the general case $n_k = n$ as follows.

$$(n-1)\tilde{S}\rho^3 - (n-2)R\rho^2 - (n\tilde{S} + S)\rho + nR = 0 \quad (\text{A.9.23})$$

with:

$$\begin{cases} S = S_{11} + \dots + S_{nn}, \\ \tilde{S} = S_{22} + \dots + S_{n-1,n-1}, \\ R = S_{12} + S_{23} + \dots + S_{n-1,n}. \end{cases}$$

Then using (A.9.18):

$$\hat{\sigma}^2 = \frac{1}{C_n} \frac{1}{(1 - \hat{\rho}^2)} (S + \hat{\rho}^2 \tilde{S} - 2\hat{\rho}R). \quad (\text{A.9.24})$$

For $n_k > 2$, (A.9.23) is a third-degree polynomial. One can show that this equation has only one root in $[-1, 1]$.

Proof. Consider:

$$\begin{aligned} f(\rho) &= (n-1)\tilde{S}\rho^3 - (n-2)R\rho^2 - (n\tilde{S} + S)\rho + nR \\ f'(\rho) &= 3(n-1)\tilde{S}\rho^2 - 2(n-2)R\rho - (n\tilde{S} + S) \\ f''(\rho) &= 6(n-1)\tilde{S}\rho - 2(n-2)R \end{aligned}$$

The discriminant of $f'(\rho)$ is as follows:

$$\Delta_{f'(\rho)} = (n-2)^2 R^2 + 3(n\tilde{S} + S)(n-1)\tilde{S} \geq 0.$$

Therefore $f'(\rho)$ has no root and hence $f(\rho)$ is monotone. One may see $f'(0) \leq 0$, therefore, $f(\rho)$ is a monotonically decreasing function (I). One can show $f(1) \leq 0$ and $f(-1) \geq 0$ (II). Considering (I) and (II) together, one may conclude $f(\rho)$ must necessarily cross the horizontal line only once between $[-1, 1]$. \square

This shows the unique $\hat{\rho}$ can be easily estimated solving (A.9.23) using Cardano's formula (Franci and Rigatelli, 1979).

A.9.3 Hessians, covariance matrices, and optimal weights

Given the MLEs for the AR(1) covariance structure, the Hessians and covariance matrices of the MLEs can be derived. Following the general results obtained about optimal weights, they can be used to compute the exact optimal weights in the case of the AR(1) structure. As mean and variance parameters are orthogonal in the normal case, we can

consider the second derivative for fixed effects and variance components separately.

Second derivative with respect to fixed effects

As

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{C_k} X_i' \Sigma^{-1} (y_i - \mu_i),$$

we have:

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell}{\partial \beta} \left(\frac{\partial \ell}{\partial \beta} \right)' \right] &= \sum_{i=1}^{C_k} X_i' \Sigma^{-1} \mathbb{E} (y_i - \mu_i) (y_i - \mu_i)' \Sigma^{-1} X_i \\ &= \sum_{i=1}^{C_k} X_i' \Sigma^{-1} X_i. \end{aligned}$$

For the special case of just an intercept $X_i = \mathbf{1}$:

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell}{\partial \beta} \left(\frac{\partial \ell}{\partial \beta} \right)' \right] &= \sum_{i=1}^{C_k} \mathbf{1}' \Sigma^{-1} \mathbf{1} = \\ &= \frac{C_k}{\sigma^2(1 - \rho^2)} \left[(n_k - 2)\rho^2 - 2(n_k - 1)\rho + n_k \right]. \quad (\text{A.9.25}) \end{aligned}$$

Therefore, the variance for $\hat{\mu}$ can be computed as inverse of (A.9.25).

Second derivative with respect to variance components

To calculate the derivatives with respect to variance components rather than $\frac{\partial C^{-1}}{\partial \rho}$, we need $K = \frac{\partial C^{-1}}{\partial \rho^2}$. Using these derivatives:

$$\begin{cases} \frac{\partial}{\partial \rho} 2 \left(\frac{\rho}{1-\rho^2} \right) = 2 \frac{1+\rho^2}{(1-\rho^2)^2}, \\ \frac{\partial}{\partial \rho} \frac{\rho}{(1-\rho^2)^2} = \frac{1+3\rho^2}{(1-\rho^2)^3}, \\ \frac{\partial}{\partial \rho} \frac{1+\rho^2}{(1-\rho^2)^2} = \frac{2\rho(3+\rho^2)}{(1-\rho^2)^3}. \end{cases} \quad (\text{A.9.26})$$

it follows that

$$\frac{\partial C^{-1}}{\partial \rho^2} = K = \frac{1}{(1-\rho^2)^3} \begin{pmatrix} 2(1+3\rho^2) & -2\rho(3+\rho^2) & & 0 \\ -2\rho(3+\rho^2) & 4(1+3\rho^2) & \ddots & \ddots \\ & \ddots & \ddots & \ddots \\ 0 & & -2\rho(3+\rho^2) & 2(1+3\rho^2) \end{pmatrix} \quad (\text{A.9.27})$$

The second-derivatives are:

$$\begin{cases} \frac{\partial^2 \ell}{\partial (\sigma^2)^2} = \frac{C_k n_k}{2} \frac{1}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^{C_k} (y_i - \mu_i)' C^{-1} (y_i - \mu_i), \\ \frac{\partial^2 \ell}{\partial \rho^2} = \frac{C_k (n_k - 1)(1+\rho^2)}{(1-\rho^2)^2} - \frac{1}{2\sigma^2} \sum_{i=1}^{C_k} (y_i - \mu_i)' K (y_i - \mu_i), \\ \frac{\partial^2 \ell}{\partial \rho \partial \sigma^2} = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{C_k} (y_i - \mu_i)' \frac{\partial C^{-1}}{\partial \rho} (y_i - \mu_i). \end{cases} \quad (\text{A.9.28})$$

To construct the expected Hessian and covariance matrix, one needs to find the expectations of the expressions in (A.9.28).

$$\mathbb{E} \left(\frac{\partial^2 \ell}{\partial (\sigma^2)^2} \right) = -\frac{C_k n_k}{2} \frac{1}{(\sigma^2)^2}. \quad (\text{A.9.29})$$

This follows from the fact that:

$$\mathbb{E} \left(\sum_{i=1}^{C_k} (y_i - \mu_i)' C^{-1} (y_i - \mu_i) \right) = C_k \text{tr} \left\{ \mathbb{E} [(y_i - \mu_i)' (y_i - \mu_i)] C^{-1} \right\},$$

and $\mathbb{E} [(y_i - \mu_i)(y_i - \mu_i)'] = \sigma^2 C$.

For the second derivative with respect to ρ :

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \rho^2} \right] = \frac{C_k (n_k - 1) (1 + \rho^2)}{(1 - \rho^2)^2} - \frac{C_k}{2} \text{tr} \{ K S \}. \quad (\text{A.9.30})$$

Likewise:

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \rho \partial \sigma^2} \right] = \frac{C_k}{2\sigma^2} \text{tr} \left\{ C \frac{\partial C^{-1}}{\partial \rho} \right\}. \quad (\text{A.9.31})$$

Substituting for $\text{tr} \{ K S \}$ and $\text{tr} \left\{ C \frac{\partial C^{-1}}{\partial \rho} \right\}$ we get:

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \rho \partial \sigma^2} \right] = \frac{C_k (n_k - 1)}{\sigma^2} \frac{\rho}{1 - \rho^2}. \quad (\text{A.9.32})$$

$$\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \rho^2} \right] = -C_k (n_k - 1) \frac{1 + \rho^2}{(1 - \rho^2)^2}. \quad (\text{A.9.33})$$

Using (A.9.29), (A.9.32), and (A.9.33) one obtains the 2×2 Hessian matrix as follows:

$$H = -C_k \begin{pmatrix} \frac{n_k}{2(\sigma^2)^2} & -\frac{n_k-1}{\sigma^2} \frac{\rho}{1-\rho^2} \\ -\frac{n_k-1}{\sigma^2} \frac{\rho}{1-\rho^2} & (n_k-1) \frac{1+\rho^2}{(1-\rho^2)^2} \end{pmatrix}. \quad (\text{A.9.34})$$

The determinant of the Hessian in (A.9.34) is as follows:

$$\det(H) = \frac{C_k^2 (n_k-1)(n_k-(n_k-2)\rho^2)}{2(\sigma^2)^2(1-\rho^2)^2}. \quad (\text{A.9.35})$$

So,

$$-H^{-1} = \frac{1}{C_k(n_k-(n_k-2)\rho^2)} \begin{pmatrix} 2(\sigma^2)^2(1+\rho^2) & 2\rho\sigma^2(1-\rho^2) \\ 2\rho\sigma^2(1-\rho^2) & \frac{n_k}{n_k-1}(1-\rho^2)^2 \end{pmatrix}. \quad (\text{A.9.36})$$

The Hessian for the unbiased estimator differs slightly from its MLE counterpart:

$$\tilde{H} = -C_k \begin{pmatrix} \frac{n_k-1}{2(\sigma^2)^2} & -\frac{n_k-1}{\sigma^2} \frac{\rho}{1-\rho^2} \\ -\frac{n_k-1}{\sigma^2} \frac{\rho}{1-\rho^2} & (n_k-1) \frac{1+\rho^2}{(1-\rho^2)^2} \end{pmatrix}, \quad (\text{A.9.37})$$

$$\det(\tilde{H}) = \frac{C_k^2 (n_k-1)^2}{2(\sigma^2)^2(1-\rho^2)}. \quad (\text{A.9.38})$$

Therefore,

$$-\tilde{H}^{-1} = \frac{1}{C_k(n_k - (n_k - 2)\rho^2)} \begin{pmatrix} 2(\sigma^2)^2(1 + \rho^2) & 2\rho\sigma^2(1 - \rho^2) \\ 2\rho\sigma^2(1 - \rho^2) & (1 - \rho^2)^2 \end{pmatrix}. \quad (\text{A.9.39})$$

Having the covariance matrix, one may easily find the optimal weights using

$$W_{opt.} = \frac{V_k^{-1}}{\sum_{i=1}^K V_i^{-1}} \quad (\text{A.9.40})$$

The variance of an estimator obtained using the optimal weights in (A.9.40) can be calculated as $\left(\sum_{i=1}^K V_i^{-1}\right)^{-1}$.

A.9.4 Proof of Proposition ??

Proof. Consider an estimator of the form:

$$\tilde{\mu}_\alpha = \frac{1}{c} \sum_{i=1}^c \sum_{j=1}^n \alpha_j Y_{ij}, \quad (\text{A.9.41})$$

for a vector of weights $\alpha = (\alpha_1, \dots, \alpha_n)'$. Because the clusters are i.i.d. it is evident that the components of α do not depend on the cluster index i . Clearly, the requirement that $E(\tilde{\mu}_\alpha) = \mu$ implies the condition

$$\sum_{j=1}^n \alpha_j = 1. \quad (\text{A.9.42})$$

An expression of the variance of $\tilde{\mu}_\alpha$ combined with this requirement

produces the objective function:

$$Q = \sigma^2 \left(\sum_{j=1}^n \alpha_j^2 + 2 \sum_{j < k} \alpha_j \alpha_k \rho^{|j-k|} \right) - \lambda \left(\sum_{j=1}^n \alpha_j - 1 \right), \quad (\text{A.9.43})$$

with λ a Lagrange multiplier. Taking the derivative of (A.9.43) w.r.t. α leads to, after rearrangement:

$$\alpha = \frac{\lambda}{2\sigma^2} C^{-1} \mathbf{1}.$$

Given that we have an explicit form for C^{-1} , it follows that

$$\alpha = \frac{\lambda}{2\sigma^2(1+\rho)} \boldsymbol{\rho}^{(1)}, \quad (\text{A.9.44})$$

with $\boldsymbol{\rho}^{(1)} = (1, 1 - \rho, \dots, 1 - \rho, 1)'$. Combining (A.9.44) with constraint (A.9.42) leads to $\lambda = 2\sigma^2(1+\rho)/[2 + (n-2)(1-\rho)]$, hence

$$\alpha = \frac{1}{[2 + (n-2)(1-\rho)]} \boldsymbol{\rho}^{(1)},$$

establishing the MLE. This completes the proof.

A.9.5 Optimal weights in case of a general mean structure $X_i^{(k)} \beta$

Cluster size specific expressions are:

$$\widehat{\beta}_k = \left(\sum_{i=1}^{c_k} X_i^{(k)'} \Sigma_k^{-1} X_i^{(k)} \right)^{-1} \left(\sum_{i=1}^{c_k} X_i^{(k)'} \Sigma_k^{-1} Y_i^{(k)} \right) \quad (\text{A.9.45})$$

and

$$\text{var}(\widehat{\beta}_k) = V_k = \left(\sum_{i=1}^{c_k} X_i^{(k)'} \Sigma_k^{-1} X_i^{(k)} \right)^{-1}. \quad (\text{A.9.46})$$

The combination rule is

$$\tilde{\beta}_k = \sum_{i=1}^K A_k \widehat{\beta}_k, \quad (\text{A.9.47})$$

with

$$V_k^{-1} = \frac{1}{\sigma^2} \sum_{i=1}^{c_k} X_i^{(k)'} C_k^{-1} X_i^{(k)} \quad (\text{A.9.48})$$

and C_k is as described in Supplementary Materials A.9.1.

The first factor in (A.9.45) can be split into three parts:

$$\begin{aligned}
 (1 - \rho^2)X_i^{(k)'} C_k^{-1} X_i^{(k)} &= X_i^{(k)'} (1 + \rho^2) I_k X_i^{(k)} \\
 &\quad - \rho^2 X_i^{(k)'} \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & & & \vdots \\ \vdots & & \ddots & & \\ \vdots & & & 0 & 0 \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix} X_i^{(k)} \\
 &\quad - \rho X_i^{(k)'} \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix} X_i^{(k)}.
 \end{aligned}$$

(A.9.5) simplifies to $(1 + \rho^2)X_i^{(k)'} X_i^{(k)}$, while (A.9.5) equals

$$\rho^2 \begin{pmatrix} x_{ki11}^2 + x_{kin_k1}^2 & x_{ki11} \cdot x_{ki22} + x_{kin_k1} \cdot x_{kin_k2} & \dots & x_{ki11} \cdot x_{ki1p} + x_{kin_k1} \cdot x_{kin_kp} \\ & x_{ki12}^2 + x_{kin_k2}^2 & & \\ & & \ddots & \\ & & & x_{ki1p}^2 + x_{kin_kp}^2 \end{pmatrix}. \quad (\text{A.9.49})$$

Defining $\mathbf{X}_{i1}^{(k)} = (x_{ki11} \dots x_{ki1p})^t$ and $\mathbf{X}_{in_k}^{(k)} = (x_{kin_k1} \dots x_{kin_kp})^t$,

(A.9.5) equals

$$\rho^2[\mathbf{X}_{i1}^{(k)} \ 0 \dots 0 \ \mathbf{X}_{in_k}^{(k)}]X_i^{(k)}. \quad (\text{A.9.50})$$

For the third term, define $\mathbf{X}_{ij}^{(k)-} = (x_{ki2j} \ \dots \ x_{kin_kj})^t$ and $\mathbf{X}_{ij}^{(k)+} = (x_{ki1j} \ \dots \ x_{kin_{k-1}j})^t$. As a consequence, (A.9.5) will equal

$$-\rho[X_i^{(k)-'} \cdot X_i^{(k)+} + X_i^{(k)+'} \cdot X_i^{(k)-}]. \quad (\text{A.9.51})$$

In summary:

$$\begin{aligned} (1 - \rho^2)X_i^{(k)'} C_k^{-1} X_i^{(k)} &= (1 + \rho^2)X_i^{(k)'} X_i^{(k)} \\ &\quad - \rho^2[\mathbf{X}_{i1}^{(k)} \ 0 \dots 0 \ \mathbf{X}_{in_k}^{(k)}]X_i^{(k)} \\ &\quad - \rho[X_i^{(k)-'} \cdot X_i^{(k)+} + X_i^{(k)+'} \cdot X_i^{(k)-}] \\ &\stackrel{\text{notation}}{=} F_{1k}. \end{aligned} \quad (\text{A.9.52})$$

The second factor in (A.9.45), using the same notations for $Y_i^{(k)}$ as

described above, can be rewritten as:

$$\begin{aligned}
 (1 - \rho^2)X_i^{(k)'} C_k^{-1} Y_i^{(k)} &= (1 + \rho^2)X_i^{(k)'} Y_i^{(k)} \\
 &\quad - \rho^2 \begin{pmatrix} x_{ki11} \cdot y_{ki1} + x_{kin_k1} \cdot y_{kin_k} \\ x_{ki12} \cdot y_{ki1} + x_{kin_k2} \cdot y_{kin_k} \\ \vdots \\ x_{ki1p} \cdot y_{ki1} + x_{kin_kp} \cdot y_{kin_k} \end{pmatrix} \\
 &\quad - \rho[X_i^{(k)-'} \cdot Y_i^{(k)+} + X_i^{(k)+'} \cdot Y_i^{(k)-}] \\
 &\stackrel{\text{not.}}{=} F_{2k}.
 \end{aligned} \tag{A.9.53}$$

Combining (A.9.52) and (A.9.53) the overall estimate equals:

$$\begin{aligned}
 \tilde{\beta}_k &= \sum_{i=1}^K A_k \widehat{\beta}_k \\
 &= \sum_{i=1}^K \left(\sum_{m=1}^K F_{1m} \right)^{-1} F_{2k}
 \end{aligned} \tag{A.9.54}$$

A.9.6 Delta method for the mean estimator

By plugging in ρ_k and defining $a'_k = c_k[n_k - (n_k - 2)\rho_k]$, equation (A.9.55) simplifies to

$$a_k = \frac{a'_k}{\sum_{m=1}^K a'_m}, \tag{A.9.55}$$

and (??) becomes

$$\text{var}(\hat{\mu}_k) = v_k = \frac{\sigma_k^2(1 + \rho_k)}{a'_k}. \quad (\text{A.9.56})$$

The first derivatives equal

$$\begin{aligned} \frac{\partial \tilde{\mu}}{\partial \mu_k} &= a_k = \frac{a'_k}{\sum_{m=1}^K a'_m}, \\ \frac{\partial \tilde{\mu}}{\partial \sigma_k^2} &= 0, \\ \frac{\partial \tilde{\mu}}{\partial \rho_k} &= \frac{-c_k(n_k - 2) \sum_{m=1}^K a'_m(\mu_k - \mu_m)}{\left(\sum_{m=1}^K a'_m\right)^2}, \end{aligned} \quad (\text{A.9.57})$$

and these can be combined using the delta method, resulting in (??):

$$\begin{aligned} \text{var}(\tilde{\mu}) &= \sum_{i=1}^K \frac{a_k'^2}{\left(\sum_{k=1}^K a'_k\right)^2} \cdot \frac{\sigma_k^2(1 - \rho_k^2)}{a'_k} \\ &+ \frac{\sum_{k=1}^K \left[c_k(n_k - 2) \sum_{m=1}^K a'_m(\mu_k - \mu_m) \right]^2}{\left(\sum_{k=1}^K a'_k\right)^4} \cdot \frac{1 - \rho_k^2}{c_k(n_k - 1)}. \end{aligned}$$

A.9.7 Calculating $\hat{\rho}$ and $\hat{\sigma}^2$ in R

In this section, we consider the implementation of the calculations for the variance components via the R package **fastAR1**. This can be done with a few simple lines of code. For fixed $C_k = C$ and $n_k = n$, and given the data y , the function `est.ar1` estimates the variance components and provides a plot for the third-degree polynomial in (A.9.23). This

visually underscores that there is only one root in $[-1, 1]$. Figure A.9.14 shows (A.9.23) for 10 simulated data sets; clearly, there is a single root only in $[-1, 1]$. For convenience, the R code can be obtained via R package `fastAR1`, see Section ?? . Other functions to find variances and iterated optimal weights are also available.

A.10 Details on additional simulations

A.10.1 Simulations with proportional and size-proportional weights

Here we consider $C_1 = 500, C_2 = 250, C_3 = 250, C_4 = 500$, and $n_1 = 5, n_2 = 10, n_3 = 10, n_4 = 5$. Parameters are set to $\mu = 0, \sigma = 2$ and $\rho = (0.1, 0.5, 0.8)$. The data are generated 100 times and the model is fitted using PROC MIXED in SAS (for a single overall intercept). For combining the results from different sub-samples we have used proportional weights and size-proportional weights:

$$\begin{cases} \text{Prop} = \frac{C_k}{\sum_l C_l}, \\ \text{Size.Prop} = \frac{C_k n_k}{\sum_l C_l n_l}. \end{cases} \quad (\text{A.10.1})$$

The results are compared with full likelihood (Table A.10.13). In contrast to the compound-symmetry case, the size-proportional weights show much better results than the proportional weights. Furthermore, the size-proportional weights in the current simulation are identical with the equal weights. The n_k 's have a much larger influence in the AR(1)

Table A.10.13: Comparing proportional, size-proportional and iterated optimal weights with full likelihood for AR(1) covariance structure.

		$\hat{\mu}$	Sd	$\hat{\rho}$	Sd	$\hat{\sigma}^2$	Sd
$\rho = 0.1$	Prop.	-0.00190	0.01615	0.10027	0.01158	1.99642	0.03002
	Size.Prop.	-0.00207	0.01538	0.10024	0.01080	1.99793	0.02853
	It.Opt.	-0.00206	0.01538	0.10024	0.00993	1.99792	0.02850
	ML	-0.00207	0.01538	0.10032	0.01078	1.99793	0.02850
$\rho = 0.5$	Prop.	-0.00212	0.02221	0.49966	0.00954	2.00349	0.03652
	Size.Prop.	-0.00155	0.02156	0.49955	0.00898	2.00257	0.03494
	It.Opt.	-0.00168	0.02149	0.49956	0.00826	2.00265	0.03486
	ML	-0.00170	0.02150	0.49986	0.00896	2.00259	0.03488
$\rho = 0.8$	Prop.	0.00195	0.02890	0.79923	0.00549	1.99529	0.04989
	Size.Prop.	0.00234	0.02904	0.79911	0.00530	1.99542	0.04907
	It.Opt.	0.00213	0.02855	0.79915	0.00486	1.99538	0.04859
	ML	0.00212	0.02855	0.79937	0.00527	1.99519	0.04861

case compared to CS. Figures A.10.15, A.10.16, and A.10.17 make the comparisons easier.

A.10.2 Simulations with proportional and size-proportional weights: ρ near 0/1

We now present a comparison between proportional and size-proportional weights. We see that, for ρ 's near 1 (i.e., near CS), size-proportional weights are worse than proportional weights.

We consider $c_1 = 500$, $c_2 = 250$, $c_3 = 250$, $c_4 = 500$, and $n_1 = 5$, $n_2 = 10$, $n_3 = 10$, $n_4 = 5$. Parameters are set as $\mu = 0$, $\sigma = 2$ and $\rho \in \{0.01, 0.2, 0.5, 0.8, 0.9, 0.95, 0.99\}$. The data are generated 100 times and the model is fitted using PROC MIXED in SAS (for a single overall intercept). For combining results from different sub-samples we have

used proportional weights and size-proportional weights as in (A.10.1). The results are compared with full likelihood results.

In Figure A.10.18, for $\rho = 0.99$ and 0.95 , the size-proportional weights perform worse than the proportional weights. This is expected, because in this case AR(1) approaches CS. This result is clearer in the left panel of Figure A.10.18, where the standard deviations are shown. For ρ 's near 1, the proportional weights are as efficient as full likelihood, while as ρ moves further from 1 this would happen for size-proportional weights.

Figure A.10.19 shows this phenomenon more clearly, as for some selected ρ 's (0.01, 0.5, 0.95) the density plot for all 100 simulated datasets is plotted rather than a boxplot. The size-proportional weights are better than proportional weights if ρ is not very close to 1. As soon as ρ becomes 0.95 or 0.99, the size-proportional weights become worse.

A.10.3 Simulations with optimal weights

Given the covariance matrix of the parameter estimators, finding the optimal weights is straightforward, but in practice the unknown parameters therein need to be estimated. Here we compare the iterative weights with size-proportional weights and ML. See Figures A.10.20, A.10.21, A.10.22, and Table A.10.13. As expected, the optimal weights lead to estimates very close to the MLE; the difference between them being numerical.

Size-proportional weights are used as the initial weights to begin the

iterative procedure. One interesting outcome of this simulation is that the iterative procedure always converged after just one iteration. This means, iterated optimal weights are just like the approximated optimal weights, but there, instead of using $\hat{\theta}_k$ from each sub-sample, one may use $\tilde{\theta}$ obtained from all sub-samples using a non-optimal but good weight.

A.10.4 Simulations on computation time

Here, some summary tables are presented to summarize the results which are already presented via figures earlier. Furthermore, a table and a figure are added to compare computation time for closed form solutions to numerical ones.

In each table the mean of the estimated parameter and its standard deviation using the 100 replications are given, together with the standard deviation of those 100 numbers (in parentheses). If θ is the parameter of interest, $\hat{\theta}$ is its estimate and θ_0 is its real value, then the MSE is computed as follows:

$$\text{MSE}(\hat{\theta}) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta} - \theta_0)^2. \quad (\text{A.10.2})$$

Table A.10.14 summarizes the results for μ . The sample splitting estimates are computed using proportional and size-proportional (identical to equal weights in this case) weights. The results using the full sample are also given. The third column in Table A.10.14 presents the averaged (over 100 replications) estimated μ and its standard deviation. The

fourth column presents the averaged estimated standard deviation for $\hat{\mu}$ (over 100 replications) and its standard deviation. The last column shows the MSE computed using (A.10.2) for $\mu_0 = 1$. Tables A.10.15 and A.10.16 shows the same results for ρ and σ^2 ($\sigma_0^2 = 2$), respectively.

Table A.10.17 compares the computation time between closed-form and iterative methods. The closed-form solutions are implemented in R and for the numerical methods the MIXED procedure in SAS is used, with error covariance structure set to AR(1).

The data are generated using $n = 10$ for all clusters, with c is varying from 100 to 1000000. Therefore, the design is balanced and the point of this comparison is to see how the computation time is reduced in each split.

As one may see in Table A.10.17 and Figure A.10.23, using closed form solutions significantly reduces the computation time. This means, as well as the computation time reduction due to splitting the data, using closed form solutions within each split the computation time reduction is also huge: for example, for a million clusters, the reduction is from almost one hour to less than 5 seconds. Figure A.10.23 shows that computation time using closed form solution changes linearly with the number of clusters, while this will be exponential using an iterative solution.

To assess the effect of the overall size of the dataset, the model is fitted to two concatenated copies of the same set. Computation time results are

presented in Table A.10.17 and Figure A.10.23. The data are generated with $\mu = 0$, $\sigma^2 = 2$, and $\rho = 0.25$.

Table A.10.14: Simulation study. Estimating μ and its standard deviation. The mean (standard deviation) of the 100 replications are given together with mean squared errors for $\rho = 0.01, 0.2, 0.5, 0.8, 0.9, 0.99$ using proportional and size-proportional weights comparing with the full likelihood results.

ρ_0	method	mean($\hat{\mu}$) (s.d.)	mean(s.e.($\hat{\mu}$)) (s.d.)	MSE $\times 10^4$
0.01	Prop.	-0.00271 (0.01462)	0.01503 (0.00020)	2.19000
	Size Prop.	-0.00277 (0.01320)	0.01429 (0.00018)	1.80172
	Full	-0.00275 (0.01319)	0.01428 (0.00018)	1.79828
0.2	Prop.	0.00158 (0.01677)	0.01752 (0.00025)	2.81056
	Size Prop.	0.00085 (0.01616)	0.01673 (0.00021)	2.59147
	Full	0.00090 (0.01615)	0.01672 (0.00021)	2.58880
0.5	Prop.	0.00391 (0.02244)	0.02217 (0.00037)	5.13770
	Size Prop.	0.00397 (0.02191)	0.02155 (0.00035)	4.91201
	Full	0.00396 (0.02182)	0.02148 (0.00034)	4.87038
0.8	Prop.	0.00130 (0.02790)	0.02894 (0.00050)	7.72450
	Size Prop.	-0.00049 (0.02710)	0.02912 (0.00045)	7.27464
	Full	0.00053 (0.02713)	0.02862 (0.00045)	7.29130
0.9	Prop.	-0.00828 (0.03006)	0.03221 (0.00056)	9.63393
	Size Prop.	-0.00727 (0.03145)	0.03306 (0.00070)	10.3224
	Full	-0.00803 (0.02998)	0.03207 (0.00057)	9.54258
0.99	Prop.	0.00162 (0.03663)	0.03597 (0.00064)	13.3123e
	Size Prop.	-0.00007 (0.03930)	0.03797 (0.00088)	15.2876
	Full	0.00156 (0.03666)	0.03597 (0.00064)	13.3305

A.11 Details on PANSS data analysis

As one may see from Table A.11.18, by far the majority of the study subjects have complete data and hence belong to the first pattern.

Figure A.11.24 presents boxplots for the entire set of data, for the

Table A.10.15: Simulation study. Estimating ρ and its standard deviation. The mean (standard deviation) of the 100 replications are given together with mean squared errors for $\rho = 0.01, 0.2, 0.5, 0.8, 0.9, 0.99$ using proportional and size-proportional weights comparing with the full likelihood results.

ρ_0	method	mean($\hat{\rho}$) (s.e.)	mean(s.e.($\hat{\rho}$)) (s.e.)	MSE
0.01	Prop.	0.01077 (0.01237)	0.01165 (0.00006)	1.52178e-04
	Size Prop.	0.01115 (0.01200)	0.01087 (0.00004)	1.43974e-04
	Full	0.01123 (0.01203)	0.01084 (0.00004)	1.44675e-04
0.2	Prop.	0.19960 (0.01213)	0.01133 (0.00007)	1.45806e-04
	Size Prop.	0.19973 (0.01145)	0.01058 (0.00005)	1.29974e-04
	Full	0.19986 (0.01142)	0.01056 (0.00005)	1.29174e-04
0.5	Prop.	0.49956 (0.00963)	0.00954 (0.00011)	9.19119e-05
	Size Prop.	0.49965 (0.00904)	0.00898 (0.00008)	8.11057e-05
	Full	0.49990 (0.00904)	0.00896 (0.00008)	8.08877e-05
0.8	Prop.	0.79973 (0.00541)	0.00548 (0.00012)	2.90660e-05
	Size Prop.	0.79990 (0.00483)	0.00529 (0.00009)	2.31132e-05
	Full	0.80018 (0.00489)	0.00525 (0.00009)	2.36726e-05
0.9	Prop.	0.90017 (0.00286)	0.00321 (0.00008)	8.14862e-06
	Size Prop.	0.90013 (0.00297)	0.00318 (0.00008)	8.76257e-06
	Full	0.90040 (0.00292)	0.00312 (0.00008)	8.60114e-06
0.99	Prop.	0.98994 (0.00038)	0.00039 (0.00001)	1.45292e-07
	Size Prop.	0.98992 (0.00042)	0.00041 (0.00002)	1.77848e-07
	Full	0.98997 (0.00037)	0.00039 (0.00001)	1.37289e-07

subjects from the first pattern only, and for various split samples.

To examine the choice of an AR(1) covariance structure, Table A.11.19 shows three model selection criteria to compare different error covariance structures. Changing from independence structure ($R = \sigma^2 I$) to compound-symmetry ($R = \sigma^2 I$) the criteria decrease with a large amount, and the same when changing to AR(1). The step to an unstructured covariance does not make a big difference (considering that the

Table A.10.16: Simulation study. Estimating σ^2 and its standard deviation. The mean (standard deviation) of the 100 replications are given together with mean squared errors for $\rho = 0.01, 0.2, 0.5, 0.8, 0.9, 0.99$ using proportional and size-proportional weights comparing with the full likelihood results.

ρ_0	method	mean($\hat{\sigma}^2$) (s.e.)	mean(s.e.($\hat{\sigma}^2$)) (s.e.)	MSE
0.01	Prop.	1.99964 (0.02960)	0.02981 (0.00049)	8.67280e-04
	Size Prop.	2.00165 (0.02836)	0.02834 (0.00040)	7.98842e-04
	Full	2.00167 (0.02832)	0.02832 (0.00040)	7.97002e-04
0.2	Prop.	2.00581 (0.02907)	0.03093 (0.00055)	8.70077e-04
	Size Prop.	2.00484 (0.02778)	0.02936 (0.00044)	7.87298e-04
	Full	2.00473 (0.02772)	0.02933 (0.00044)	7.83248e-04
0.5	Prop.	1.99783 (0.03860)	0.03638 (0.00097)	1.47960e-03
	Size Prop.	1.99890 (0.03747)	0.03488 (0.00085)	1.39116e-03
	Full	1.99897 (0.03748)	0.03481 (0.00085)	1.39152e-03
0.8	Prop.	2.00013 (0.04900)	0.05002 (0.00166)	2.37744e-03
	Size Prop.	2.00136 (0.04423)	0.04930 (0.00137)	1.93881e-03
	Full	2.00101 (0.04569)	0.04880 (0.00142)	2.06754e-03
0.9	Prop.	2.00122 (0.05767)	0.05872 (0.00196)	3.29407e-03
	Size Prop.	2.00089 (0.06037)	0.05915 (0.00230)	3.60829e-03
	Full	2.00115 (0.05876)	0.05793 (0.00198)	3.41986e-03
0.99	Prop.	1.99683 (0.06941)	0.07117 (0.00254)	4.77911e-03
	Size Prop.	1.99527 (0.07598)	0.07484 (0.00344)	5.73813e-03
	Full	1.99641 (0.06940)	0.07093 (0.00253)	4.78099e-03

unstructured covariance would have 21 parameters to estimate compared to 2 parameters in the AR(1) model). Therefore, AR(1) seems to be a good choice.

The 95% confidence intervals, accompanying (??), are presented in Figure A.11.25. In order to give more insight in these results, Figure A.11.26 shows the 95% confidence interval in each split, comparing with the full sample splits (the horizontal dashed line in the figure).

Table A.10.17: Simulation study. The computation time for a sample with $n = 10$ and $c = 1\text{e}+02, 1\text{e}+03, 1\text{e}+04, 5\text{e}+04, 1\text{e}+05, 3\text{e}+05, 5\text{e}+05, 7\text{e}+05, 9\text{e}+05, 1\text{e}+06$. The closed form solution is obtained by implementing the results of this paper in R, and the numerical solution is obtained using PROC MIXED in SAS to estimate a repeated measurement model with AR(1) covariance structure.

time (s)	1e+02	1e+03	1e+04	5e+04	1e+05	3e+05	5e+05	7e+05	9e+05	1e+06
Closed form	0.00	0.00	0.03	0.23	0.34	1.45	2.07	3.37	4.40	4.89
Numerical	0.08	0.13	1.04	10.45	34.74	268.96	770.74	1611.43	2724.31	3399.47

The 95% confidence intervals, accompanying (??), are presented in Figure A.11.27. Figure A.11.28 shows the 95% confidence intervals for the parameter estimates in each split comparing with the full sample estimate (the horizontal dashed-like in the figure.)

A.12 Fieller’s method and delta method

Here, the Fieller’s method to construct a confidence interval for (??) as well as the Delta method are discussed. The ratio which we require a confidence interval for is:

$$n_f = \frac{\hat{\sigma}^2}{\widehat{\tau\varepsilon}}.$$

Suppose n_1 is a small-feasible sub-sampling size chosen to estimate the parameters in the model (e.g., $n_1 = 5$). From Molenberghs et al. (2011), one finds:

Table A.11.18: PANSS data. Number of clusters in each trial for each cluster pattern.

n	Pattern	Trial					Total
		FIN-1	FRA-3	INT-2	INT-3	INT-7	
2	* *	17	8	71	43	3	142
	* . *	0	0	2	0	1	3
	* . . . * . .	0	0	1	0	0	1
3	* * *	8	4	83	41	7	143
	* . * . * . .	0	0	2	0	0	2
	* * . . * . .	1	0	3	1	0	5
4	* * * . * . .	11	0	85	66	5	167
	* . * . * . *	0	0	1	0	1	2
	* . * . * . *	0	0	1	0	0	1
	* * * . . . *	0	0	3	0	0	3
	* * * * . . .	0	4	1	0	0	5
	* * . * . . *	0	1	0	0	0	1
	* . * . . . *	0	0	0	0	1	1
	* * * . * . *	0	0	0	0	0	0
5	* * * . * . *	58	0	85	35	6	184
	* * * . * . *	0	0	8	0	1	9
	* * . . . * *	0	0	6	0	0	6
	* * * . . . *	0	0	8	0	0	8
	* . * . * . *	0	0	3	0	2	5
	* . * . * . *	0	0	1	0	0	1
	* * * * . . .	0	44	0	0	0	44
	* * . * * *	0	1	0	0	0	1
6	* * * . * . *	0	0	986	240	74	1300
	* * . . . * *	0	0	1	0	0	1
	* * * . . . *	0	0	1	0	0	1
	* * * . * . *	0	0	1	0	0	1
	* * * . * * *	0	0	2	0	0	2
	* * * . * * *	0	0	2	0	0	2

$$\begin{aligned}
 \text{Var} \begin{pmatrix} \widehat{\sigma^2} \\ \widehat{\tau\varepsilon} \end{pmatrix} &= \\
 &= \frac{2\sigma^4}{Nn_1(n_1-1)} \begin{pmatrix} n_1 & -\varepsilon \\ -\varepsilon & \varepsilon^2 \frac{\sigma^4 + 2(n_1-1)\tau\sigma^2 + n_1(n_1-1)\tau^2}{\sigma^4} \end{pmatrix} \\
 &= \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}. \quad (\text{A.12.1})
 \end{aligned}$$

Table A.11.19: PANSS data. Comparing different error covariance structures using three model comparison criteria for model (??) (residual log-likelihood value; AIC; BIC). Three R structures: Ind. : independence structure ($R = \sigma^2 I$), CS: compound-symmetry structure ($R = \sigma^2 I + dJ$), AR(1): AR(1) structure ($R_{ij} = \sigma^2 \rho^{|i-j|}$), UN: unstructured ($R_{ij} = \sigma_{ij}^2$).

Model	-2 Res.log.lik.	AIC	BIC
Unstructured	80005.1	80047.1	80164.1
AR(1)	80522.6	80526.6	80537.8
Compound symm.	82683.1	82687.1	82698.3
Independence	89546.1	89548.1	89553.7

The Fieller's confidence interval for (??) can be calculated following these steps.

$$\begin{aligned}
 C_1^2 &= \frac{s_{11}}{\widehat{\sigma^2}^2}, \\
 C_2^2 &= \frac{s_{22}}{(\varepsilon\widehat{\tau})^2}, \\
 r &= \frac{s_{12}}{\sqrt{s_{11}s_{22}}}, \\
 A &= C_1^2 + C_2^2 - 2rC_1C_2, \\
 B &= t^2 C_1^2 C_2^2 (1 - r^2), \\
 L &= \frac{\widehat{\sigma^2}}{\varepsilon\widehat{\tau}} \frac{1 - z_{\alpha/2}^2 r C_1 C_2 - z_{\alpha/2} \sqrt{A - B}}{1 - z_{\alpha/2}^2 C_2^2}, \\
 U &= \frac{\widehat{\sigma^2}}{\varepsilon\widehat{\tau}} \frac{1 - z_{\alpha/2}^2 r C_1 C_2 + z_{\alpha/2} \sqrt{A - B}}{1 - z_{\alpha/2}^2 C_2^2}.
 \end{aligned} \tag{A.12.2}$$

One may take $n_f = \max\{L, U\}$. Of course, in case were $z_{\alpha/2}^2 C_2^2 < 1$, there would be no finite interval. As it was mentioned, one may also use

the Delta method as follows:

$$\text{Var}\left(\frac{\widehat{\sigma^2}}{\varepsilon^2 \widehat{\tau^2}}\right) \approx \frac{1}{\varepsilon^2 \widehat{\tau^2}^2} \left(s_{11} - 2 \frac{\sigma^2}{\tau} s_{12} + \frac{\sigma^4}{\tau^2} s_{22} \right). \quad (\text{A.12.3})$$

A.13 Random vertical data splitting for compound-symmetry structure

Assume a single cluster of size N with multivariate normal distribution and compound-symmetry structure for its covariance matrix, see Section ???. Suppose we take M_0 sub-samples of size n_0 from this cluster. Following Hoffman et al. (2001), the parameter μ should be estimated within each sub-sample, $\widehat{\mu}^{(m)}$ ($m = 1, \dots, M_0$) and then the overall estimate can be obtained by averaging these estimates. Furthermore, one needs the within and between sub-samples variabilities to compute the overall variability, see (??).

Within each sub-sample the variance of $\widehat{\mu}^{(m)}$ can be computed using the variance formula for a mean estimator in CS-multivariate normal Molenberghs et al. (2011) :

$$S_W = \frac{\sigma^2}{n_0} [1 + (n_0 - 1)\rho]. \quad (\text{A.13.1})$$

Estimating the between variability requires computing $E\{(\widehat{\mu}^{(m)} - \bar{\mu})^2\}$, where $\bar{\mu}$ is the average of $\widehat{\mu}^{(m)}$'s ($m = 1, \dots, M_0$). Set S_m as the index

of members of sub-sample m and $S_{m'}$ the rest of the sample, then,

$$\hat{\mu}^{(m)} - \bar{\mu} = \frac{1}{n_0} \left\{ \left(1 - \frac{1}{M_0}\right) \sum_{i \in S_m} y_i - \frac{1}{M_0} \sum_{m=1}^{M_0} \sum_{i \in S_{m'}} y_i \right\}.$$

Therefore,

$$\mathbb{E} \left\{ (\hat{\mu}^{(m)} - \bar{\mu})^2 \right\} = \frac{1}{n_0} (T_1 + T_2 + T_3 + T_4), \quad (\text{A.13.2})$$

where,

$$\begin{cases} T_1 = \left(1 - \frac{1}{M_0}\right) \mathbb{E} [\sum_{i \in S_m} y_i]^2 \\ T_2 = \frac{1}{M_0^2} (M_0 - 1) \mathbb{E} [\sum_{i \in S_m} y_i]^2 \\ T_3 = -2 \left(1 - \frac{1}{M_0}\right) \frac{1}{M_0} (M_0 - 1) \mathbb{E} \left\{ \left(\sum_{i \in S_m} y_i \right) \left(\sum_{i' \in S_{m'}} y_{i'} \right) \right\} \\ T_4 = 2 \frac{1}{M_0} \frac{(M_0 - 1)(M_0 - 2)}{2} \mathbb{E} \left\{ \left(\sum_{i \in S_m} y_i \right) \left(\sum_{i' \in S_{m'}} y_{i'} \right) \right\}. \end{cases} \quad (\text{A.13.3})$$

From (A.13.3) one may find,

$$\begin{cases} T_1 + T_2 = \frac{M_0 - 1}{M_0} n_0 \sigma^2 [1 + (n_0 - 1)\rho] \\ T_3 + T_4 = -\frac{M_0 - 1}{M_0} \frac{\sigma^2 n_0^2}{N} [1 + (N_1)\rho]. \end{cases} \quad (\text{A.13.4})$$

Therefore, the between variability can be computed as follows,

$$S_B = \mathbb{E} \left\{ (\hat{\mu}^{(m)} - \bar{\mu})^2 \right\} = \frac{M_0 - 1}{M_0} \left[\frac{1 + (n_0 - 1)\rho}{n_0} - \frac{1 + (N_1)\rho}{N} \right]. \quad (\text{A.13.5})$$

Using (A.13.1) and (A.13.5), the total variability can be computed as follows,

$$S_T = \frac{1}{M_0} \sigma^2 \frac{1 + (n_0 - 1)\rho}{n_0} + \frac{M_0 - 1}{M_0} \sigma^2 \frac{1 + (N - 1)\rho}{N} \quad (\text{A.13.6})$$

Comparing the variance in (A.13.6) with the variance when estimating μ using the full sample would give,

$$\text{ARE} = \frac{M_0 - 1}{M_0} + \frac{1}{M_0} \frac{N}{n_0} \frac{1 + (n_0 - 1)\rho}{1 + (N - 1)\rho} \quad (\text{A.13.7})$$

For example, if sub-sampling is done only once ($M_0 = 1$), then,

$$\text{ARE} = \frac{N}{n_0} \frac{1 + (n_0 - 1)\rho}{1 + (N - 1)\rho}, \quad (\text{A.13.8})$$

The same calculation is possible for $M_0 = 2$,

$$\text{ARE} = \frac{1}{2} + \frac{1}{2} \frac{N}{n_0} \frac{1 + (n_0 - 1)\rho}{1 + (N - 1)\rho}. \quad (\text{A.13.9})$$

By considering the desired ARE as $1 + \epsilon$, one may find n_0 .

A.14 Combination rule for p -values for MI

The Kruskal-Wallis test statistic asymptotically follows a χ^2 distribution. Therefore, in order to find the combined p -value, we may follow the procedure proposed by Li et al. (1991). More details can be found in Rubin (2004) and Enders (2010). The combination is done using `micombine.chisquare` in package `miceadds` in R, Robitzsch et al.

(2016).

- **Averaging.** Compute the test statistic for each imputed data and take their average: $\overline{\chi^2} = \frac{1}{M} \sum_{m=1}^M \chi_m^2$.
- **Relative variance increase.** $r = \left(1 + \frac{1}{M}\right) \frac{\sum_{m=1}^M (\sqrt{\chi_m^2} - \sqrt{\overline{\chi^2}})^2}{M-1}$
- **Test statistic.** $D = \frac{\overline{\chi^2} - \frac{M+1}{M-1}r}{1+r}$, where κ is the degrees-of-freedom of the χ_m^2 . For a χ^2 -test of independence, it is $(k_1-1)(k_2-1)$ where k_1 and k_2 are the number of levels of two categorical variables. Take five diagnosis groups and variable **Eye** with two levels, then $\kappa = (5-1)(2-1) = 4$. For the Kruskal-Wallis test, κ is the number of groups minus 1. So, when comparing values of a continuous variable among three types of diagnosis, $\kappa = 2$. When this comparison is done for all five types, $\kappa = 4$.
- **Computing p -value.** $p\text{-value} = P(F_{\kappa, \nu} > D)$, where F is the F -distribution and $\nu = \kappa^{-3/M}(M-1)\left(1 + \frac{1}{r}\right)^2$

A.15 The surrogate model

The model takes the form:

$$\begin{cases} S_{ij} = \mu_S + m_{S_i} + (\alpha + a_i)Z_{ij} + \varepsilon_{S_{ij}}, \\ T_{ij} = \mu_T + m_{T_i} + (\beta + b_i)Z_{ij} + \varepsilon_{T_{ij}}, \end{cases} \quad (\text{A.15.1})$$

where S_{ij} and T_{ij} are the surrogate and true endpoints, and Z_{ij} is the treatment indicator (1 for active; 0 for placebo), respectively. Index i refers to the trial and j to the subject. Further, μ_S and μ_T are fixed intercepts and m_{S_i} and m_{T_i} are random intercepts, for surrogate and true endpoint, respectively. The fixed treatment effects are α (for S) and β (for T), and the centre-specific treatment effects are a_i (for S) and b_i (for T). Random intercepts and slopes are assumed to follow $(m_{S_i}, m_{T_i}, a_i, b_i)' \sim N(0, D)$, where

$$D = \begin{pmatrix} d_{SS} & & & \\ d_{ST} & d_{TT} & & \\ d_{Sa} & d_{Ta} & d_{aa} & \\ d_{Sb} & d_{Tb} & d_{ab} & d_{bb} \end{pmatrix}. \quad (\text{A.15.2})$$

Furthermore, for the error terms,

$$\begin{pmatrix} \varepsilon_{S_{ij}} \\ \varepsilon_{T_{ij}} \end{pmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_{SS} & \\ \sigma_{ST} & \sigma_{TT} \end{pmatrix}. \quad (\text{A.15.3})$$

The parameters of interest for the surrogate evaluation are trial- and individual-level coefficients of determination, R_{trial}^2 and R_{indiv}^2 , respectively:

$$R_{\text{trial}}^2 = \frac{1}{d_{ab}} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}, \quad (\text{A.15.4})$$

$$R_{\text{indiv}}^2 = \frac{\sigma_{ST}^2}{\sigma_{SS}\sigma_{TT}}.$$

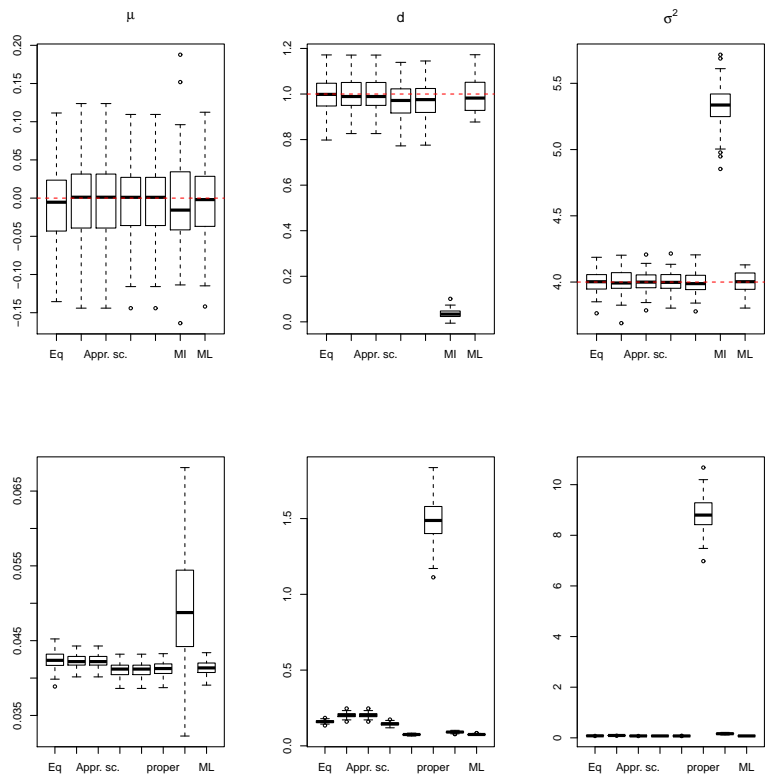


Figure A.7.13: Second simulation study. Estimated CS parameters (first row) and their standard error (second row) using sample splitting, MI-MLE, and MLE.

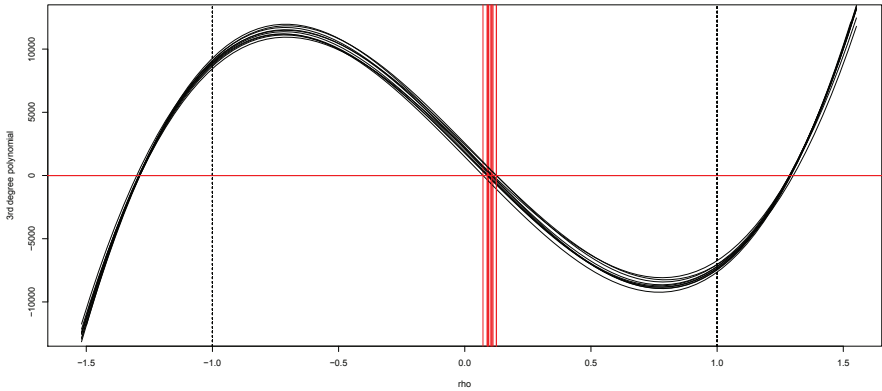


Figure A.9.14: The third degree polynomial in (A.9.23) for 10 different generated data. The red vertical line shows $\hat{\rho}$.

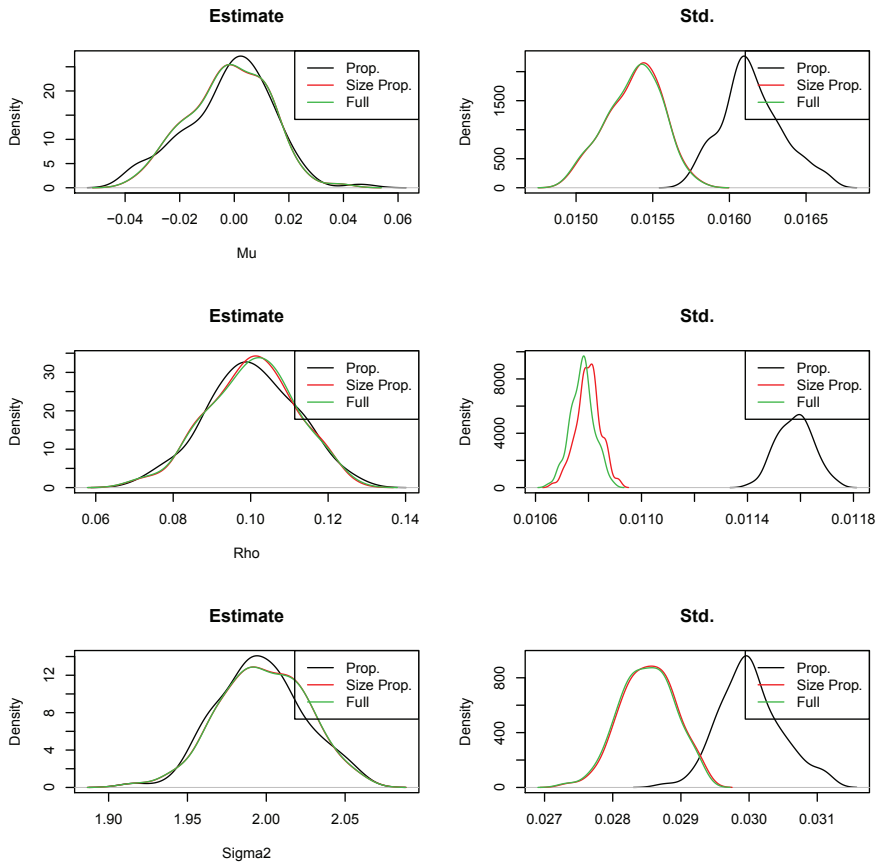


Figure A.10.15: Comparing proportional and size-proportional weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.1$.

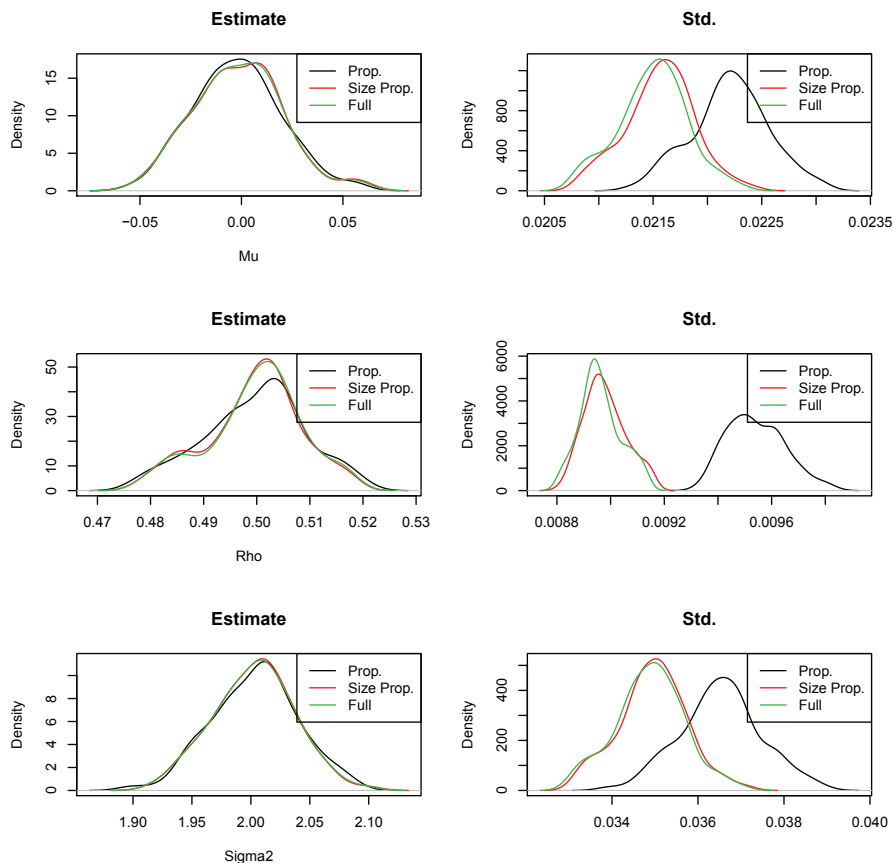


Figure A.10.16: Comparing proportional and size-proportional weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.5$.

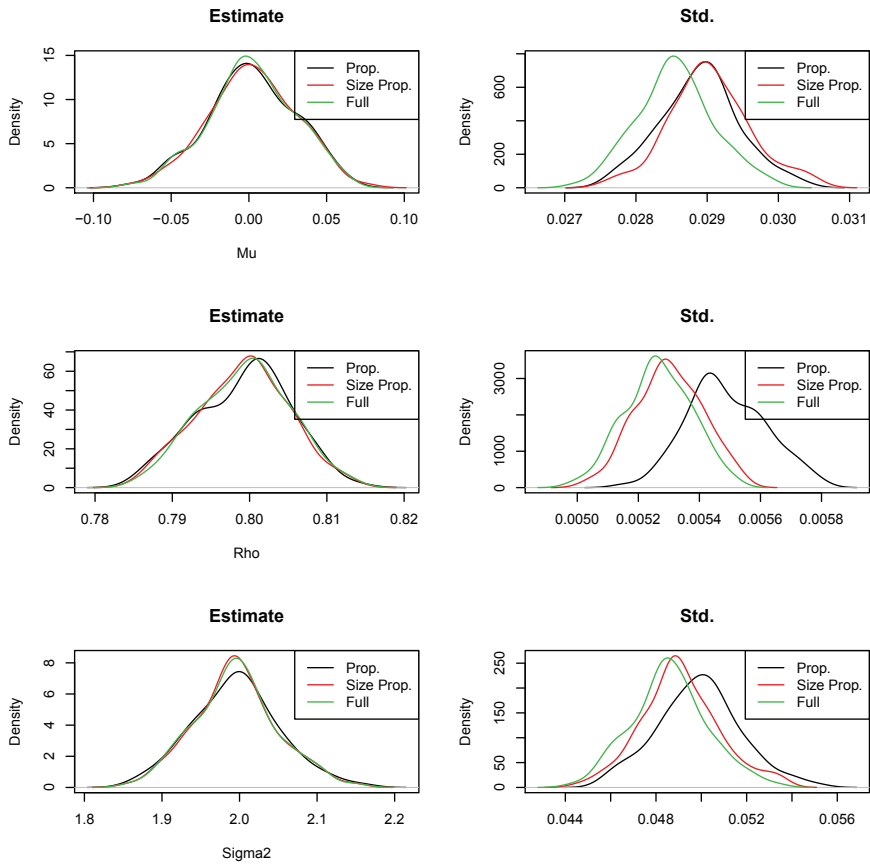


Figure A.10.17: Comparing proportional and size-proportional weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.8$.

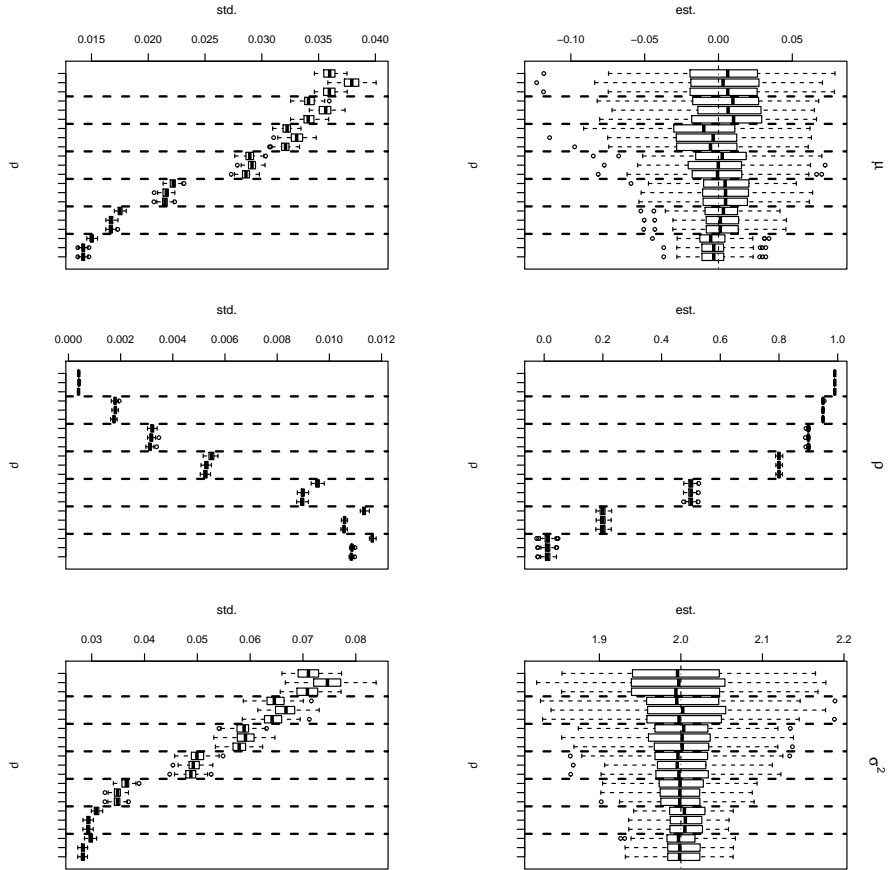


Figure A.10.18: Simulation study. Boxplots comparing proportional and size-proportional weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.99, 0.95, 0.9, 0.8, 0.5, 0.2, 0.01$. In every section of the boxplots (which are separated by dashed lines) the first out of three represents the proportional weights, the middle of is size-proportional weights and the one on the right shows the results for the full likelihood. The first row presents the estimates while the second row shows the standard deviations of these estimates.

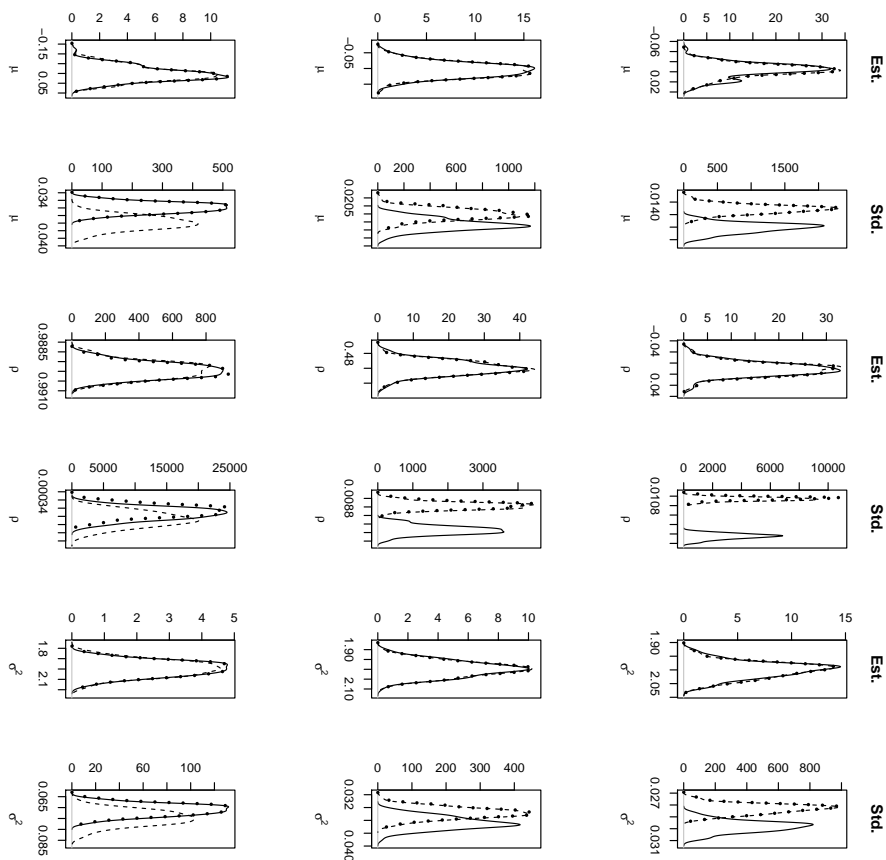


Figure A.10.19: Simulation study. Comparing proportional, size-proportional and full likelihood results via their empirical density for the 100 replications. In all of the figures $\mu = 0$ and $\sigma^2 = 2$. The first row is for $\rho = 0.01$, the middle one is for $\rho = 0.5$ and last one corresponds to $\rho = 0.99$. In each figure the ticked dotted line corresponds to full likelihood, the dashed line is for size-proportional weights and the solid line is for proportional weights.

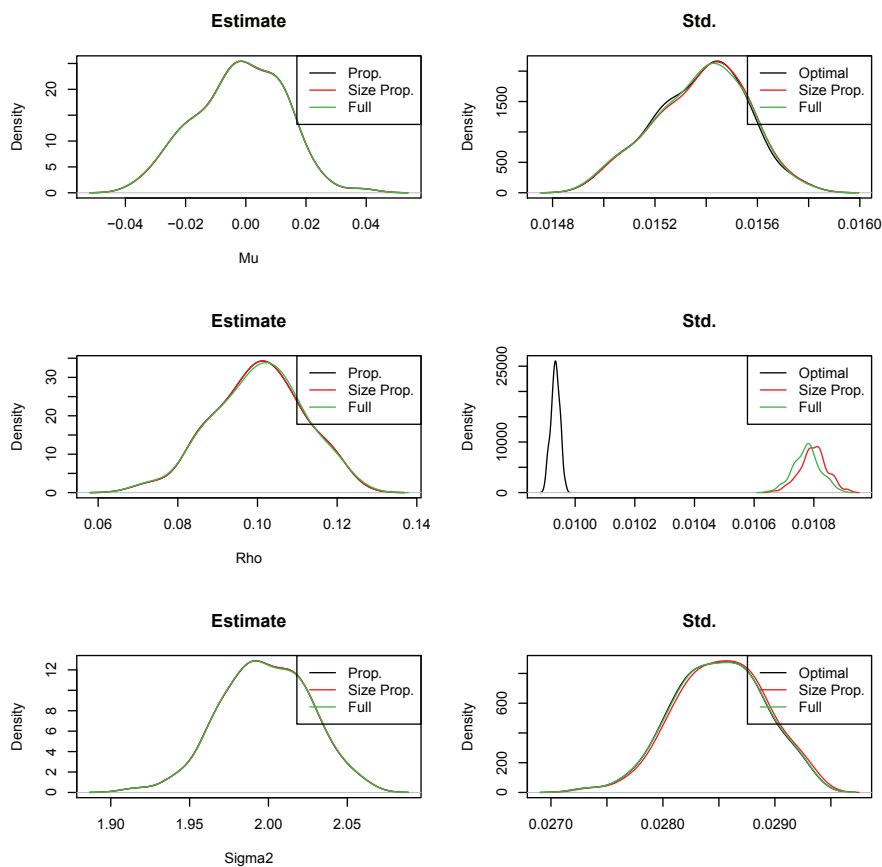


Figure A.10.20: Comparing iterated optimal and size-proportionanl weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.1$.

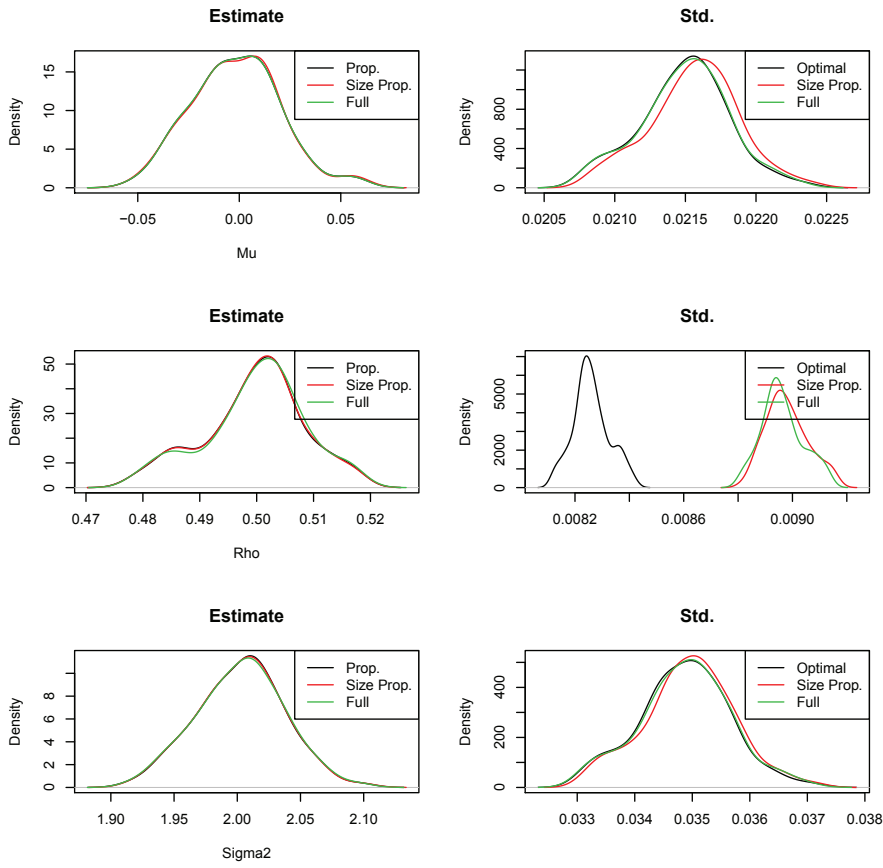


Figure A.10.21: Comparing iterated optimal and size-proportionanl weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.5$.

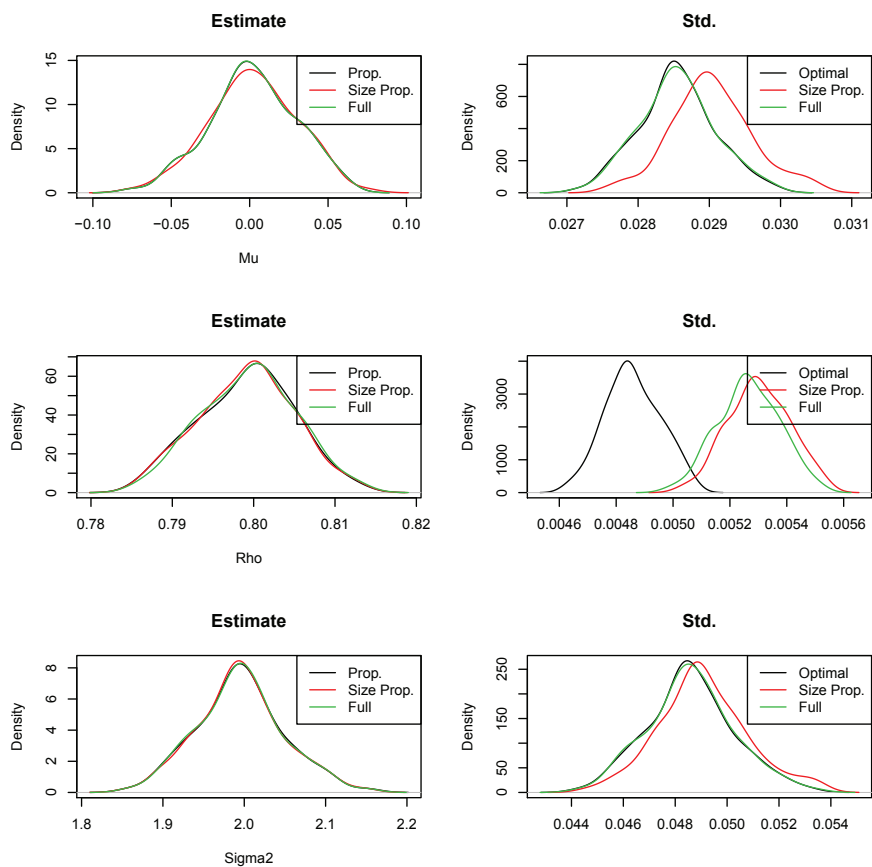


Figure A.10.22: Comparing iterated optimal and size-proportional weights with full likelihood for 100 replications with $\mu = 0$, $\sigma^2 = 2$ and $\rho = 0.8$.

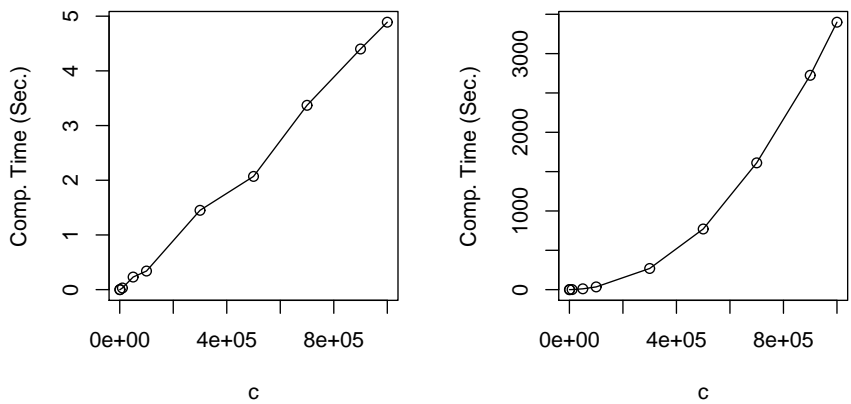


Figure A.10.23: Simulation study. Comparing computation time using closed form (left) and numerical (right) solutions. The horizontal axis shows number of clusters (c) and the vertical axis shows the computation time in seconds.

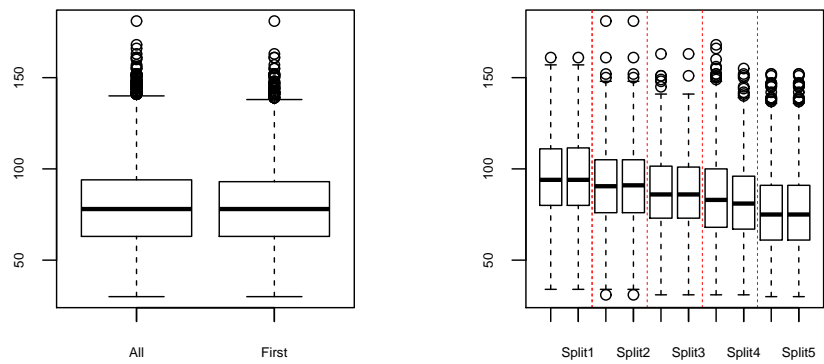


Figure A.11.24: PANSS data. Boxplots for the entire set of data, for the subject from the first pattern only, and for various split samples.

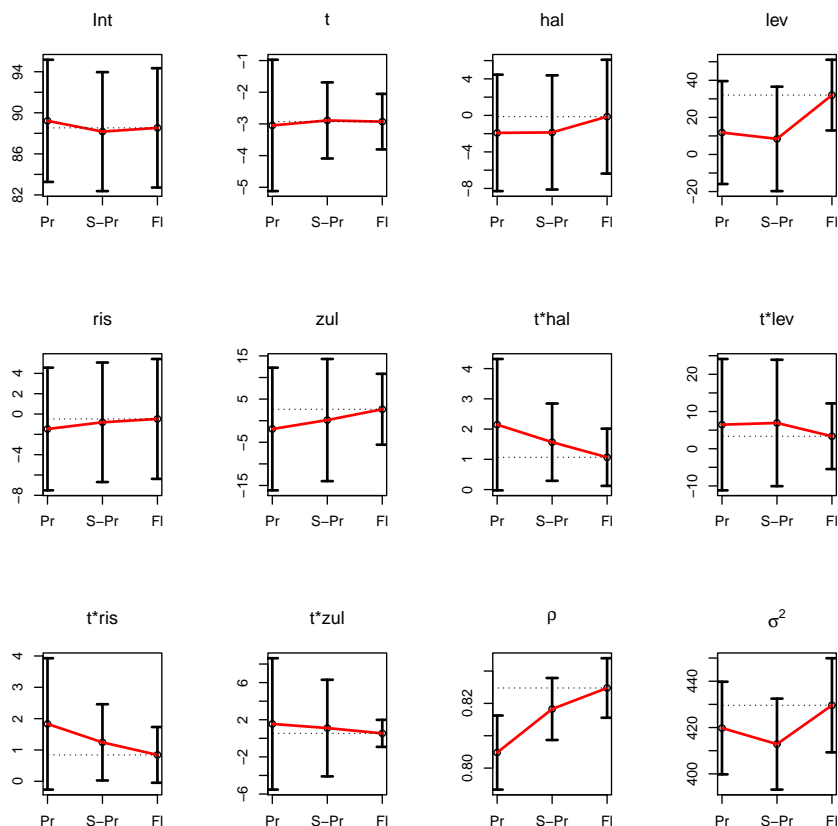


Figure A.11.25: PANSS data. 95% confidence intervals for fixed effects and variance components estimates and the standard deviations of these estimates using sample splitting, combined with proportional (Pr - first) and size-proportional (S-Pr - second) weights, and full likelihood (FI - third). The dashed horizontal line shows the full likelihood estimate. The model used in here is without trial effect (??).

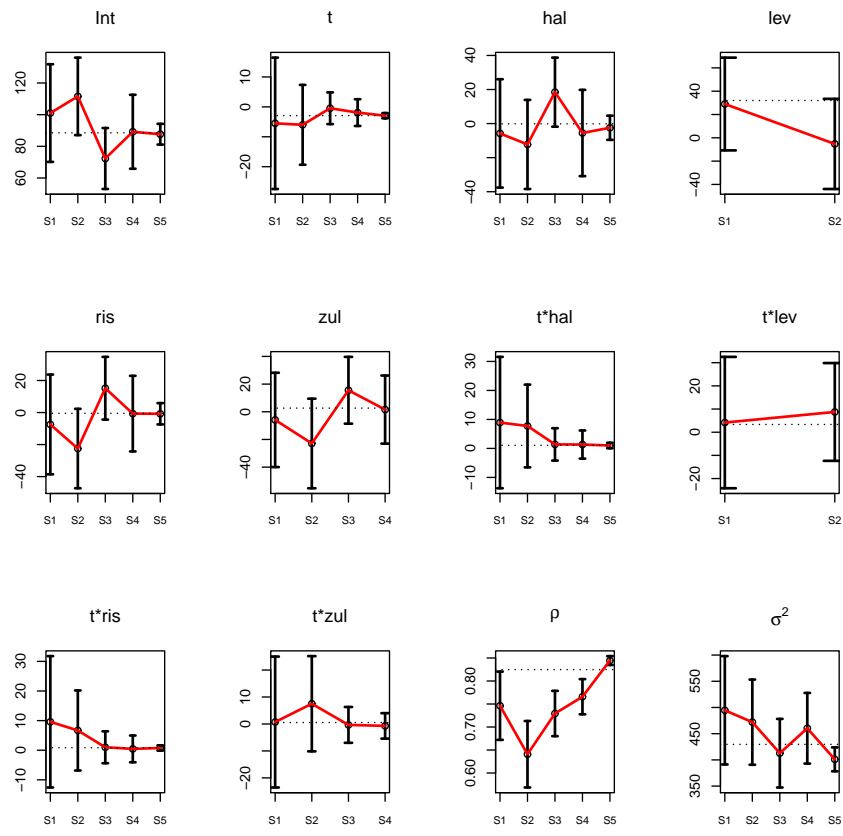


Figure A.11.26: PANSS data. 95% confidence intervals for fixed effects and variance components estimates and the standard deviations of these estimates within each split. The dashed horizontal line shows the full likelihood estimate. The model used in here is without trial effect (??).

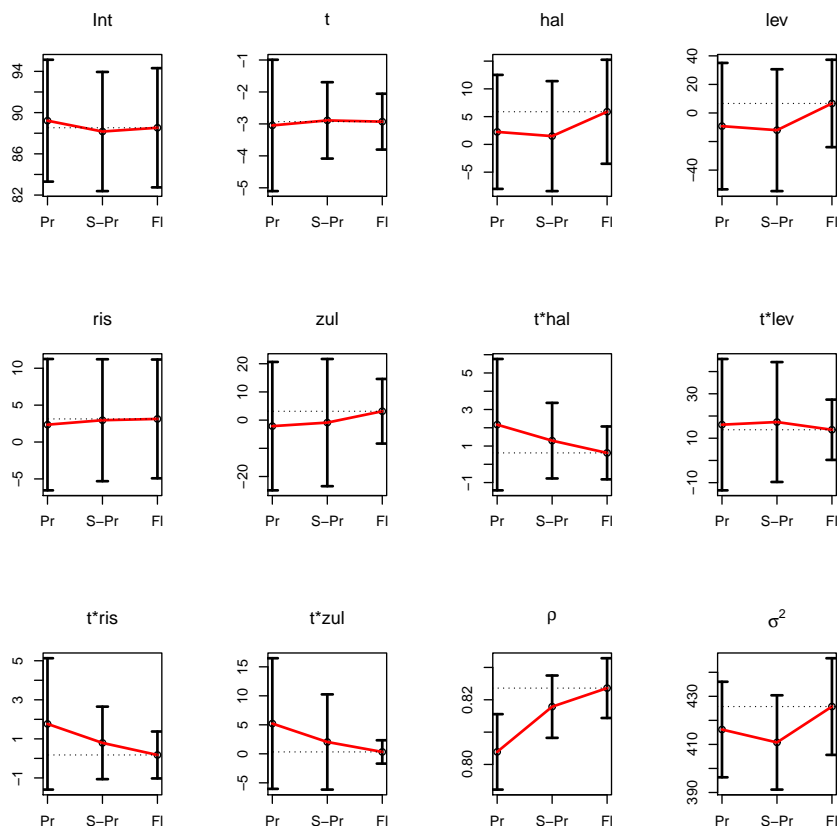


Figure A.11.27: PANSS data. 95% confidence intervals for fixed effects and variance components estimates and the standard deviations of these estimates using sample splitting, combined with proportional (Pr - first) and size-proportional (S-Pr - second) weights, and full likelihood (FI - third). The dashed horizontal line shows the full likelihood estimate. The model used in here is with trial effect (??).

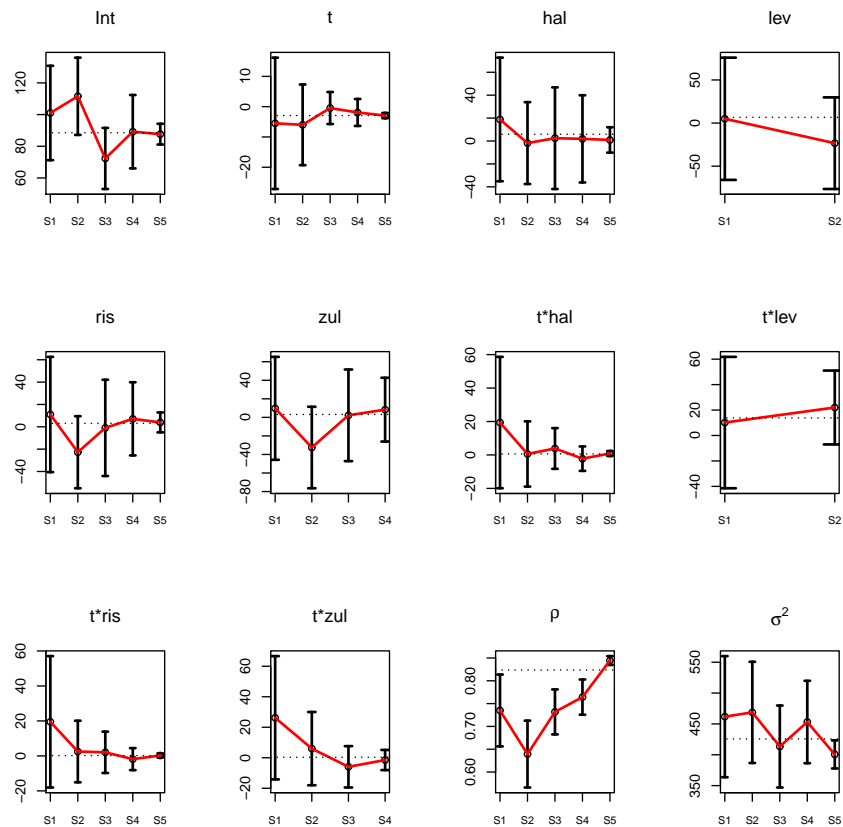


Figure A.11.28: PANSS data. 95% confidence intervals for fixed effects and variance components estimates and the standard deviations of these estimates within each split. The dashed horizontal line shows the full likelihood estimate. The model used in here is with trial effect (??).

BIBLIOGRAPHY

B. C. Arnold and D. Strauss. Pseudolikelihood estimation: some examples. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 233–243, 1991.

C. K. Enders. *Applied missing data analysis*. Guilford Press, 2010.

S. Fieuws and G. Verbeke. "pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles". *Biometrics*, 62(2):424–431, 2006. ISSN 0006-341X.

S. Fieuws, G. Verbeke, F. Boen, and C. Delecluse. High dimensional multivariate mixed models for binary questionnaire data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(4): 449–460, 2006.

- R. Franci and L. T. Rigatelli. *Storia della teoria delle equazioni algebriche*, volume 40. Ugo Mursia Editore, 1979.
- E. B. Hoffman, P. K. Sen, and C. R. Weinberg. Within-cluster resampling. *Biometrika*, 88:1121–1134, 2001.
- N. M. Laird, J. H. Ware, et al. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
- N. Lange and N. M. Laird. The effect of covariance structure on variance estimation in balanced growth-curve models with random parameters. *Journal of the American Statistical Association*, 84(405):241–247, 1989.
- K.-H. Li, X.-L. Meng, T. E. Raghunathan, and D. B. Rubin. "Significance levels from repeated p -values with multiply-imputed data". *Statistica Sinica*, pages 65–92, 1991.
- G. Molenberghs and G. Verbeke. *Models for discrete longitudinal data*. 2005.
- G. Molenberghs, G. Verbeke, and S. Iddi. "Pseudo-likelihood methodology for partitioned large and complex samples". *Statist. Probab. Lett.*, 81:892–901, 2011. ISSN 0167-7152.
- W. Neiswanger, C. Wang, and E. Xing. Asymptotically exact, embarrassingly parallel mcmc. *arXiv preprint arXiv:1311.4780*, 2013.
- A. Robitzsch, S. Grund, and T. Henke. Miceadds: Some additional multiple imputation functions, especially for 'mice', 2016. URL <https://cran.r-project.org/web/packages/miceadds/index.html>.

- D. B. Rubin. "The design of a general and flexible system for handling nonresponse in sample surveys". *The American Statistician*, 58:298–302, 2004.
- S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch. Bayes and big data: The consensus monte carlo algorithm. *International Journal of Management Science and Engineering Management*, 11(2):78–88, 2016.
- K. Sikorska, E. Lesaffre, P. F. Groenen, and P. H. Eilers. Gwas on your notebook: fast semi-parallel linear and logistic regression for genome-wide association studies. *BMC bioinformatics*, 14(1):166, 2013.
- G. Verbeke and S. Fieuws. The effect of miss-specified baseline characteristics on inference for longitudinal trends in linear mixed models. *Biostatistics*, 8(4):772–783, 2007.
- G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data*. Springer Series in Statistics. Springer-Verlag, New York, 2000.