Linear Models for Correlated Data: Inference

The goal is to estimate the vector of regression coefficients $\boldsymbol{\beta}$ when the data are correlated. We assume

$$\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, V)$$
$$\boldsymbol{Y}_i \sim MVN(X_i\boldsymbol{\beta}, V_i),\ i = 1, \ldots, m$$

where V and V_i are covariance matrices

- *Balanced data* $\Rightarrow V_i = V_0,\ i = 1, \ldots, m$
- *Unbalanced data* $\Rightarrow V_i \neq V_0,\ i = 1, \ldots, m$
- Parametric models for covariance matrix
- Completely unstructured covariance matrix

Inference

- Weighted Least Square (WLS) (V_i known)
- Maximum Likelihood (V_i unknown)
- Restricted Maximum Likelihood (V_i unknown)
- Robust estimation (V_i unknown)
- Hypothesis Testing
- Example: Growth of Sitka Trees

Weighted least-squares estimation

$$\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, V)$$

- The *weighted least squares* estimate of $\boldsymbol{\beta}$, using a symmetric weight matrix W, is the value $\tilde{\boldsymbol{\beta}}_W$, which minimizes the quadratic form:

$$(\boldsymbol{y} - X\boldsymbol{\beta})^{'} W (\boldsymbol{y} - X\boldsymbol{\beta})$$

- the solution is:

$$\tilde{\boldsymbol{\beta}}_W = (X^{'}WX)^{-1}X^{'}W\boldsymbol{y}$$

- **$\tilde{\boldsymbol{\beta}}_W$ is an unbiased estimator of β whatever the choice of W**

Weighted least-squares estimation

- If $W = \sigma^2 I$ then $\tilde{\boldsymbol{\beta}}_W = \tilde{\boldsymbol{\beta}}_I$, where $\tilde{\boldsymbol{\beta}}_I$ is the ordinary least-squares estimator

$$\tilde{\boldsymbol{\beta}}_I = (X^{'}X)^{-1}X^{'}\boldsymbol{y}$$

- $var(\tilde{\boldsymbol{\beta}}_I) = \sigma^2(X^{'}X)^{-1}$
- If $W = V^{-1}$ and $\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, V)$ then $\tilde{\boldsymbol{\beta}}_W = \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the MLE $\boldsymbol{\beta}$ so defined:

$$\hat{\boldsymbol{\beta}} = (X^{'}V^{-1}X)^{-1}X^{'}V^{-1}\boldsymbol{y}$$

- $var(\hat{\boldsymbol{\beta}}) = (X^{'}V^{-1}X)^{-1}$
- the most efficient weighted least-squares estimator for $\boldsymbol{\beta}$ uses $W = V^{-1}$
- Why? Because by using $W = V^{-1}$, then $\hat{\boldsymbol{\beta}}$ maximizes the likelihood function

Estimation of Mean Model
Weighted Least Squares

- **General Linear Model** for longitudinal data:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$$\boldsymbol{\epsilon} \sim MVN(0, V)$$

- How the regression parameters $\boldsymbol{\beta}$ are estimated?
- The log-likelihood of $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = -\frac{1}{2}nmlog(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta})^{'}V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})$$

- Therefore the **maximum likelihood estimator** $\hat{\boldsymbol{\beta}}$ is obtained by **minimizing** the weighted sum of squares

$$WRSS = (\boldsymbol{y} - X\boldsymbol{\beta})^{'}V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})$$

- $\hat{\boldsymbol{\beta}}$ that minimizes WRSS is a **weighted least squares** with $W = V^{-1}$ and it is defined as:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X^{'}V^{-1}X)^{-1}X^{'}V^{-1}\boldsymbol{y} \\ var(\hat{\boldsymbol{\beta}}) &= (X^{'}V^{-1}X)^{-1} \end{aligned}$$

- If the data are **indipendent**, then V takes the form $V = \sigma^2 I$ which gives rise to the OLS estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{OLS} &= (X^{'}X)^{-1}X^{'}\boldsymbol{y} \\ var(\hat{\boldsymbol{\beta}}) &= \sigma^2(X^{'}X)^{-1} \end{aligned}$$

Note that we can re-write the WRRS as following:

$$\begin{aligned} WRSS &= (\boldsymbol{y} - X\boldsymbol{\beta})^{'}V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta}) \\ &= \textstyle\sum_{i=1}^{m}(\boldsymbol{y}_i - X_i\boldsymbol{\beta})^{'}V_i^{-1}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}) \\ &= \textstyle\sum_{i=1}^{m}(\boldsymbol{y}_i^{\star} - X_i^{\star}\boldsymbol{\beta})^{'}(\boldsymbol{y}_i^{\star} - X_i^{\star}\boldsymbol{\beta}) \end{aligned}$$

where:

$$\begin{aligned} \boldsymbol{y}_i^{*} &= V_i^{-\frac{1}{2}}\boldsymbol{y}_i \\ X_i^{*} &= V_i^{-\frac{1}{2}}X_i \end{aligned}$$

Therefore WLS is equivalent to OLS applied to transformed data $\boldsymbol{y}^{\star}$ and $X^{\star}$. In fact

$$\hat{\boldsymbol{\beta}} = (X^{'}V^{-1}X)^{-1}X^{'}V^{-1}\boldsymbol{y} = (X^{*'}X^{*})^{-1}(X^{*'}\boldsymbol{y}^{*})$$

.

What Does this Equation Say? Examples

- V diagonal
- V is not a diagonal matrix, $corr(Y_1, Y_2) = .9$
- V is not a diagonal matrix, AR model of order 1

Examples: V diagonal

$$Y_i = \beta_0 + \epsilon_i,\ 1 = 1, 2, 3$$

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\beta_0 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

$$\boldsymbol{\epsilon} \sim MVN(0, V)$$

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}}_{OLS} = \frac{y_1 + y_2 + y_3}{3}$$

Examples: V diagonal

$$
\begin{aligned}
V^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{10} \end{pmatrix} \\
&= \\
\hat{\boldsymbol{\beta}}_{WLS} &= (1^{'}V^{-1}1)^{-1}1^{'}V^{-1}\boldsymbol{y} \\
&= \\
&= (2.1)^{-1}(y_1 + y_2 + .1y_3) \\
&= \\
&= .48y_1 + .48y_2 + .04y_3
\end{aligned}
$$

Examples: V no diagonal

$$
\begin{aligned}
V &= \begin{pmatrix} 1 & .9 & 0 \\ .9 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
V^{-1} &= \begin{pmatrix} 5.3 & -4.7 & 0 \\ -4.7 & 5.3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
\hat{\boldsymbol{\beta}}_{WLS} &= (1^{'}V^{-1}1)^{-1}1^{'}V^{-1}\boldsymbol{y} \\
&= (2.053)^{-1}(.526y_1 + .526y_2 + .48y_3) \\
&= .26y_1 + .26y_2 + .48y_3 \\
&= .52\left(\frac{y_1+y_2}{2}\right) + .48y_3
\end{aligned}
$$

Examples: AR1

Only one subject, we assume that covariance parameters θ and σ^2 are known, and that the covariance matrix V has an exponential correlation structure

$$
\begin{aligned}
\boldsymbol{y} &= (y_1, y_2, \ldots, y_n)^{'} \\
y_j &= x_j\beta + \epsilon_j \\
\epsilon_j &= \theta\epsilon_{j-1} + a_j \\
a_j &\sim N(0, \sigma^2) \\
Cov(\epsilon_j, \epsilon_{j+\tau}) &= \sigma^2\theta^\tau \\
V &= \sigma^2 \begin{pmatrix} 1 & \theta & \theta^2 & \theta^3 & \ldots & \theta^{n-1} \\ & 1 & \theta & \theta^2 & \ldots & \theta^{n-2} \\ & & 1 & \theta & & \vdots \\ & & & 1 & & \theta^2 \\ & & & & & \theta \\ & & & & & 1 \end{pmatrix}
\end{aligned}
$$

$y_j^* = (V^{-1/2}y_j) = y_j - \theta y_{j-1},\ j = 2, \ldots, n$

$$
\begin{aligned}
y_j - \theta y_{j-1} &= x_j\beta + \epsilon_j - \theta(x_{j-1}\beta + \epsilon_{j-1}) \\
&= (x_j - \theta x_{j-1})\beta + \epsilon_j - \theta\epsilon_{j-1}
\end{aligned}
$$

where

$$
\begin{aligned}
y_j^* &= x_j^*\beta + a_j \\
a_j &\sim N(0, \sigma^2)
\end{aligned}
$$

- $y_j^* = y_j - \theta y_{j-1}$
- $x_j^* = (x_j - \theta x_{j-1})$
- $a_j = \epsilon_j - \theta\epsilon_{j-1}$

Now use OLS with y_j^* and x_j^* to get WLS estimate of β.

Weighted least-squares estimation - Summary

$$\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, V),\ V \text{ known}$$

- For an **arbitrary** W, the *weighted least squares* estimate of β is
$$\tilde{\boldsymbol{\beta}}_W = (X'WX)^{-1}X'W\boldsymbol{y}$$
- If we choose $W = V^{-1}$, then the following weighted least square estimator
$$\hat{\boldsymbol{\beta}}_W = (X'V^{-1}X)^{-1}X'V^{-1}\boldsymbol{y}$$
has minimum variance among all the weighted least squares estimators. This because $\hat{\boldsymbol{\beta}}_W$ it is also the Maximum Likelihood estimator when V is known

Efficiency

Let's assume that:

$$\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, V),\ V \text{ known}$$

- We calculate the OLS estimate assuming that the data are indipendent, i.e. $W = \sigma^2 I$:
$$\begin{aligned} \hat{\boldsymbol{\beta}}_{OLS} &= (X'X)^{-1}X'\boldsymbol{y} \\ \text{var}(\hat{\boldsymbol{\beta}}_{OLS}) &= \sigma^2(X'X)^{-1}X'VX(X'X)^{-1} \end{aligned}$$
- We do the "right thing", i.e. we calculate the WLS estimate with $W = V^{-1}$ and get the MLE:
$$\begin{aligned} \hat{\boldsymbol{\beta}}_{WLS} &= (X'V^{-1}X)^{-1}X'V^{-1}\boldsymbol{y} \\ \text{var}(\hat{\boldsymbol{\beta}}_{WLS}) &= \sigma^2(X'V^{-1}X)^{-1} \end{aligned}$$
- How bad is $\hat{\boldsymbol{\beta}}_{OLS}$ with respect to $\hat{\boldsymbol{\beta}}_{WLS}$?
- Calculate the efficiency, as ratio of the variance of the two estimators. If the ratio is close to 1, then the OLS is ok.
$$e(\hat{\boldsymbol{\beta}}_{OLS}) = \frac{\text{var}(\hat{\boldsymbol{\beta}}_{WLS})}{\text{var}(\hat{\boldsymbol{\beta}}_{OLS})}$$
- If the ratio is close to 1, then the OLS is ok.

Example

- $m = 10$ units each observed at $n = 5$ time-points $t_j = -2, -1, 0, 1, 2$
- let the mean response at time t be
$$\mu(t) = \beta_0 + \beta_1 t$$
- assume that $V_0 = (1-\rho)I + \rho\mathbf{1}\mathbf{1}^{\mathbf{T}}$
- here the OLS are fully efficient in this case:

$$\boxed{var(\tilde{\boldsymbol{\beta}}_{OLS}) = var(\hat{\boldsymbol{\beta}})}$$

where:

- $var(\tilde{\boldsymbol{\beta}}_{OLS}) = (X'X)^{-1}X'VX(X'X)^{-1}$
- $var(\hat{\boldsymbol{\beta}}) = (X'V^{-1}X)^{-1}$
- with some matrix calculations, we can show that $var(\tilde{\boldsymbol{\beta}}_{OLS}) = var(\hat{\boldsymbol{\beta}})$

When can we use OLS and ignore V?

1. uniform correlation model
2. balanced data

- with a common correlation between any **two equally spaced** measurements on the same unit there is no reason to weight measurements differently
- this would be not true if the number of measurements **varied between units** because, with $\rho > 0$, units with more measurements would then convey more information per measurements than units with fewer measurements.
- in many circumstances where there is a **balanced design**, the **OLS** estimator is perfectly **satisfactory** for point estimation.

Example: Two-treatment crossover design

- $n = 3$ measurements are taken, at unit time intervals, on each of $m = 8$ subjects
- the sequence of treatments given to the eight subjects are AAA, AAB, ABA, ABB, BAA, BAB, BBA and BBB

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$$

- where x is a binary indicator for treatment B and ϵ_{ij} follow and exponential correlation model with correlation ρ between successive measurements on any subject
- In this case, **OLS is horribly inefficient** for β when ρ is large

Example: Two-treatment crossover design

- here, efficient estimation of β_1 requires careful balancing of between-subject and within-subject comparisons of the two treatments, and the approximate balance depends critically on the correlation structure.
- In presence of positive autocorrelation, main use of ordinary least squares can seriously over or under estimate the variance of $\hat{\boldsymbol{\beta}}$, depending on the design matrix.
- here an uniform correlation model is not appropriate

So far we have developed a theory that estimate $\boldsymbol{\beta}$ in a marginal model for the mean $E[\boldsymbol{Y}] = X\boldsymbol{\beta}$, when the errors are correlated $\boldsymbol{\epsilon} \sim MVN(0, V)$, and V is known. We have learned that $\hat{\boldsymbol{\beta}} = (X'V^{-1}X)^{-1}XV^{-1}\boldsymbol{y}$ is MLE.

The problem is that we don't know V. Two options:

- If the data are balanced, $V_i = V_0$, and we are willing to assume a parametric model for V_0. In this case, we can estimate $\boldsymbol{\beta}$ and V_0 "jointly" by maximinzing the log-likelihood.
- Alternatively, we can use "robust" estimation, which does not require to specify a parametric model for V.

Maximum Likelihood estimation under Gaussian assumption

Simultaneous estimation of the parameter of interest $\boldsymbol{\beta}$ and of covariance parameters σ^2 and V_0 using the likelihood function

If $\boldsymbol{Y} \sim MVN(X\boldsymbol{\beta}, \sigma^2 V)$, the log-likelihood for observed data $\boldsymbol{y}$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2, V_0) &= -0.5\{nm\log(\sigma^2) + m\log(\mid V_0 \mid) + \\ &+ \sigma^{-2}(\boldsymbol{y} - X\boldsymbol{\beta})'V^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})\} \end{aligned}$$

1. Assume V_0 and σ^2 are known, and maximize $L(\boldsymbol{\beta}, \sigma^2, V_0)$ as function of $\boldsymbol{\beta}$. The MLE estimator for $\boldsymbol{\beta}$ is the weighted least squares estimator

$$\hat{\boldsymbol{\beta}}(V) = (X'V^{-1}X)^{-1}X'V^{-1}\boldsymbol{y}$$

2. Calculate $L(\hat{\boldsymbol{\beta}}(V), \sigma^2, V_0)$, and maximize $L(\hat{\boldsymbol{\beta}}(V), \sigma^2, V_0)$ with respect to σ^2. This gives

$$\hat{\sigma}^2(V_0) = RSS(V_0)/nm$$

where

$$RSS(V_0) = (\boldsymbol{y} - X\hat{\boldsymbol{\beta}}(V_0)^{'} V^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}(V_0))$$

3. Calculate $L(\hat{\boldsymbol{\beta}}(V), \hat{\sigma}^2, V_0)$, and maximize $L(\hat{\boldsymbol{\beta}}(V), \hat{\sigma}^2, V_0)$ with respect to V_0.

The maximum likelihood estimates are:

- $\hat{V}_0 = \text{argmax} L_r(V_0)$
- $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{V}_0)$
- $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{V}_0)$

Restricted Maximum Likelihood estimates

- MLE approach produces biased estimates of the variance components in the general linear model
- the MLE estimate of σ^2 is $\hat{\sigma}^2 = RSS/(nm)$ where RSS denotes the residual sum of squares
- an unbiased estimator for σ^2 is $\tilde{\sigma}^2 = RSS/(nm - p)$ where p denotes the number of elements of $\boldsymbol{\beta}$ - this is called Restricted Maximum Likelihood Estimator.

Generalized Least Square Estimator
Robust estimation

If we are not willing to specify a parametric model for V, then we can use a "robust" estimation and estimate $\boldsymbol{\beta}$ by:

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_W &= (X'WX)^{-1}X'W\boldsymbol{y}\\ \hat{R}_W &= \{(X'WX)^{-1}X'W\}\hat{V}\{(X'WX)^{-1}X'W\}\end{aligned}$$

where:

- $\hat{V}$ is a consistent estimate for V whatever the true covariance structure (will tell you how to calculate $\hat{V}$)
- W is a "working" covariance matrix,
- Example are: $W = I$ or $[W]_{jk} = \exp\{-c \mid t_j - t_k \mid\}$.

Then is can be show that:

$$\tilde{\boldsymbol{\beta}}_W \sim MVN(\boldsymbol{\beta}, \hat{R}_W)\ (\star)$$

Robust estimation of V under a saturated model

- measurements are made at each of n time-points t_j on m_h experimental units in g experimental groups
- $y_{hij},\ h = 1, \ldots, g,\ i = 1, \ldots, m_h,\ j = 1, \ldots, n$
- h =treatment, i =unit, and j =time-point
- the saturated model for the mean response is

$$E(Y_{hij}) = \mu_{hj},\ h = 1, \ldots, g,\ j = 1, \ldots, n$$

- a saturated model for the covariance matrix assume

$$V(Y) = V$$

with non-zero diagonal block equal to V_0,a positive definite but otherwise arbitrary $n \times n$ matrix.

Robust Estimation of V

- $\hat{\mu}_{hj} = \frac{1}{m_h}\sum_{i=1}^{m_h} y_{hij}$
- REML estimator for V_0 is:

$$\hat{V}_0 = \left(\sum_{h=1}^{g} m_h - g\right)^{-1} \times \times \sum_{h=1}^{G}\sum_{i=1}^{m_h}(\boldsymbol{y}_{hi} - \hat{\boldsymbol{\mu}}_h)(\boldsymbol{y}_{hi} - \hat{\boldsymbol{\mu}}_h)'$$

  where

  $\boldsymbol{y}_{hi} = (y_{hi1}, \ldots, y_{hin})'$

  $\boldsymbol{\mu}_h = (\mu_{h1}, \ldots, \mu_{hn})'$

- the required estimate $\hat{V}$ is the block-diagonal matrix with non zero blocks $\hat{V}_0$.

For Example

- $g = 2,\ m_1 = 2,\ m_2 = 3$ we have

$$\boldsymbol{X} = \begin{bmatrix} I & 0 \\ I & 0 \\ 0 & I \\ 0 & I \\ 0 & I \end{bmatrix}$$

  where I and O are, respectively, the $n \times n$ identity matrix and the $n \times n$ matrix of zeros.

Robust estimation versus a parametric approach

- the crucial difference between this and a parametric modeling approach is that a poor choice of W will affect only the efficiency of our inferences for $\boldsymbol{\beta}$, not their validity
- confidence intervals and test hypothesis derived from $(\star)$ will be asymptotically correct whatever the true form of V
- we can get consistent estimate of V by REML under a saturated model

Maximum Likelihood Estimation of V

When the saturated model strategy is not feasible, typically when data are from observational studies with continuous covariates, we can estimate V my maximizing the likelihood - however this depends on how big is V!.

Unbalanced Data

In this case V can still be block diagonal, but the V_{0i} will have different sizes. We can still estimate V_{0i} as:

$$\hat{V}_{0i} = (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)'$$

where $\hat{\boldsymbol{\mu}}_i$ is the OLS estimate of $\boldsymbol{\mu}_i$ from the most complicated model we are prepared to entertain for the mean response.

Example: Growth of sitka tree with and without ozone

- data consist of measurements on 79 sitka spruce trees over two growing seasons
- the trees were grown in four controlled environment chambers, of which the first two containing 27 trees each, were treated with introduced ozone at 70 ppb while the remaining two, containing 12 and 13 trees, were controls
- response variable is the log-size measurement $y = \log(hd^2)$ where h denotes height and d denoted diameter
- **Q:** is there a ozone effect on the growth pattern?
- We use a separate parameter β_j say, for the treatment mean response at the jth time-point and concentrate our modeling efforts on the control versus treatment contrast

Scatterplot matrix of residuals for the 1988 data

- You need to remove the effects of any explanatory variables, say the day and treatment
- For example, you might want to obtain the residuals from a 2-way anova model (OLS) on day and treatment group (with interaction)
- `logsize ~ day * ozone`

Example: Growth of sitka tree with and without ozone
Unstructured covariance matrix

- **Q:** Is there an effect of the ozone on the growth pattern?
- Use a saturated model for the mean, i.e.

$$E[Y_{hij}] = \mu_{hj},\ h = 1, \ldots, 4,\ j = 1, \ldots, 5(1988)$$

- We calculated the REML for V_0 in 1988 and 1989
- Chambers effects appear be negligible

Example: Growth of sitka tree with and without ozone

- Because our inferential focus is on the ozone effect, we make no attempt to model an overall growth pattern parametrically
- we assume

$$\mu_1(t_j) = \beta_j,\ j = 1, \ldots, 5$$
$$\mu_2(t_j) = \beta_j + \tau + \gamma t_j,\ j = 1, \ldots, 5$$

- we use a separate parameter, β_j, for the treatment mean response at the jth time point and concentrate the modelling effort on the control versus treatment contrast
- we estimate β_j, τ and γ by using ordinary least squares $(W = I)$

- we estimate V_0 using REML
- the hypothesis of no treatment effect is $\tau = \gamma = 0$
- test statistics $T = 9.79$ on 2 df corresponding to $p = 0.007$, i.e. strong evidence of a negative treatment effect, that is, ozone suppresses growth.

1989 Data

For the 1989 data, we assume that this contrast is linear in time, thus

$$\mu_1(t_j) = \beta_j, \; j = 1, \ldots, 5$$
$$\mu_2(t_j) = \beta_j + \tau \; j = 1, \ldots, 5$$

- the hypothesis of no treatment effect is $\tau = 0$
- test statistics T is equal to 5.15 on 1 df corresponding to $p = 0.023$.

Summary: Unstructured covariance matrix

- Robust approach here described are very simple to implement
- REML estimates of the covariance structure are simple to compute provided that the experimental design allows the fitting for a saturated model for the mean response, and the remaining calculations involve only standard matrix manipulation
- by design, consistent inferences for the mean response parameters follow from the correct specification of the mean structure, whatever the true covariance structure.

Summary: Parametric Models for covariance matrix

- Here the good reasons in favor of considering explicit modeling of the covariance structure

1. efficiency: the theoretically optimal weighted least-squares estimate uses a weight matrix whose inverse is proportional to the true covariance matrix so it would seem reasonable to use the data to estimate this optimal weight matrix
2. when there are n measurement per experimental unit, the robust approach use $\frac{1}{2}n(n+1)$ parameters to describe the covariance matrix, all of which must be estimated from the data
3. in contrast the true covariance structure may involve

mane few parameters, which can themselves be estimated more accurately than the unconstrained variance matrix

- in summary, the robust approach is usually satisfactory when the data consist of short, complete, sequences of measurements observed at a common set of times on many experimental units, and care is taken in the choice of the working correlation matrix.
- in other circumstances is worth considering a parametric modelling approach

Summary

$$\boldsymbol{Y}_i \sim MVN(X_i\boldsymbol{\beta}, \sigma^2 V_0),\ i = 1, \ldots, m$$

- V_0 known $\rightarrow$ WLS
- V_0 unknown $\rightarrow$ REML
- if V_0 is unstructured then REML can be computationally expensive
- if V_0 is unstructured $\rightarrow$ robust estimation

1. specify saturated model for the mean

$$E(Y_{hij}) = \mu_{hj}$$

2. estimate μ_{hj} by OLS and get $\hat{\mu}_{hj}$
3. REML estimate of $\hat{V}_0$
4. by using $\hat{V}_0$ get robust standard errors for $\boldsymbol{\beta}$

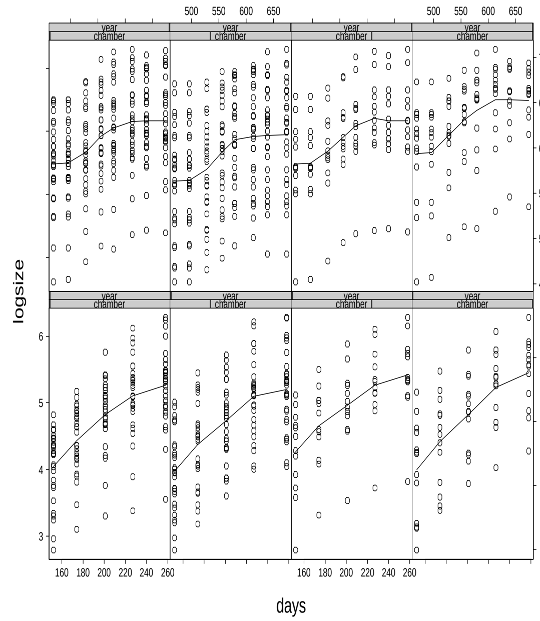

Figure 1: *Observed data and mean response profiles in each of the four growth chambers for the treatment and control.*

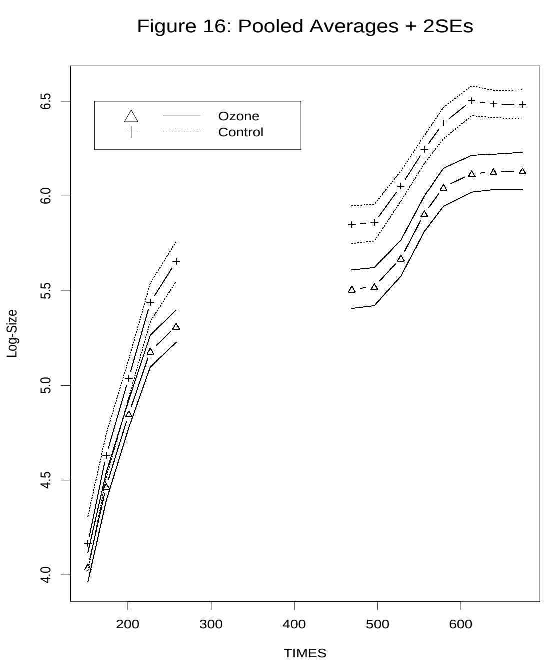

Figure 2: *Observed mean response in each of the four chambers.*

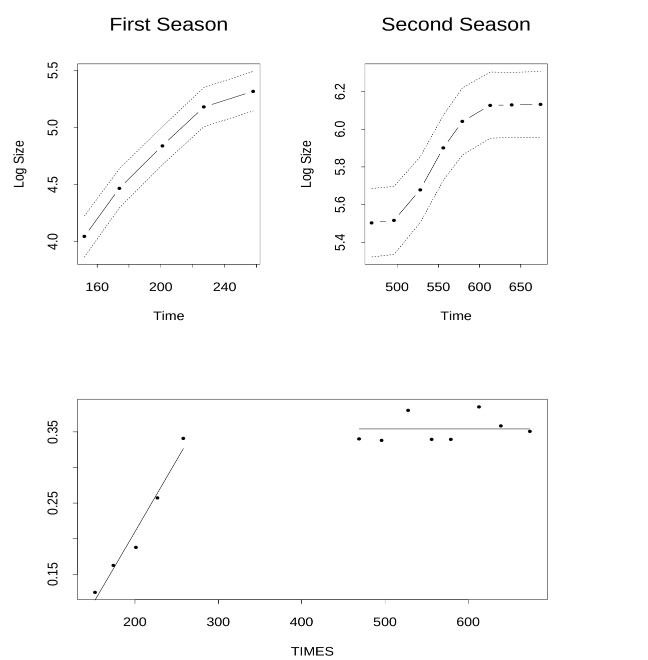

Figure 3: *Top: Estimated response profiles and 95% pointwise confidence limits. Bottom: observed and fitted differences in mean response profiles between the control and the ozone treated groups.*