Statistical Methods for Correlated Data

Assessment of Assumptions and Diagnostics (GEE and LMMs)

Michele Guindani

Department of Biostatistics
UCLA

Learning objectives

- Residuals for GEE and LMMs
- Semi-variogram as an (imperfect) diagnostic tool
- Model Selection tools for GEE

Recap Assumptions of GEE

GEE

Model:

$$\boldsymbol{Y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{e}_i,$$

with working covariance model $\text{var}(\boldsymbol{e}_i) = \boldsymbol{W}_i(\boldsymbol{\alpha})$,
$i = 1, ..., m$.

G1 Marginal model $\text{E}[\boldsymbol{Y}_i] = \boldsymbol{x}_i\boldsymbol{\beta}$ is appropriate.

G2 m is sufficiently large for asymptotic inference to be appropriate.

G3 m is sufficiently large for robust estimation of standard errors.

G4 The working covariance $\boldsymbol{W}_i(\boldsymbol{\alpha})$ is not far from the "true" covariance structure; if this is the case then the analysis will be very inefficient (standard errors will be much bigger than they need to be).

Recap Assumptions of LMER

LMEM

Model:

$$\boldsymbol{Y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with $\boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{D})$, $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{E}_i)$, $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ independent, $i = 1, ..., m$.

L1 Mean model for fixed effects $\boldsymbol{x}_i\boldsymbol{\beta}$ is appropriate.

L2 Mean model for random effects $\boldsymbol{z}_i\boldsymbol{b}_i$ is appropriate.

L3 Variance model for $\boldsymbol{\epsilon}_i$ is correct.

L4 Variance model for $\boldsymbol{b}_i$ is correct.

L5 Normality of $\boldsymbol{\epsilon}_i$.

L6 Normality of $\boldsymbol{b}_i$.

L7 m is sufficiently large for asymptotic inference to be appropriate.

Residuals

- Population-level residuals

$$\boldsymbol{e}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}$$

  are most useful for analyses based on the marginal (GEE) approach, estimated as $\widehat{\boldsymbol{e}}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}$

- Unit-level residuals

$$\boldsymbol{\epsilon}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{z}_i\boldsymbol{b}_i$$

  estimated as $\widehat{\boldsymbol{\epsilon}}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{z}_i\widehat{\boldsymbol{b}}_i$.

Residuals

- Population-level residuals

$$\boldsymbol{e}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}$$

  are most useful for analyses based on the marginal (GEE) approach, estimated as $\widehat{\boldsymbol{e}}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}$

- Unit-level residuals

$$\boldsymbol{\epsilon}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{z}_i\boldsymbol{b}_i$$

  estimated as $\widehat{\boldsymbol{\epsilon}}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{z}_i\widehat{\boldsymbol{b}}_i$.

- With dependent data, when we use the *raw* residuals above, we expect the population residuals to be dependent:

$$\text{var}\left(\hat{\boldsymbol{e}}_i\right) = \hat{V}_i - X_i\left(X_i'\hat{V}_i^{-1}X_i\right)^{-1}X'$$

  where $\hat{V}_i$ is the estimated total variance of Y_i, from either maximum likelihood estimate or the restricted maximum likelihood (REML) estimator.

- Similarly, The variance-covariance matrix of the conditional residuals can be specified as follows:

$$\text{var}\left(\hat{\boldsymbol{\varepsilon}}_i\right) = \left(\mathbf{I}_{n_i} - Z_i \hat{\mathbf{G}} Z_i' \hat{V}_i^{-1}\right) \text{var}\left(\hat{\boldsymbol{e}}_i\right) \left(\mathbf{I}_{n_i} - Z_i \hat{\mathbf{G}} Z_i' \hat{V}_i^{-1}\right)'$$

- In linear mixed models, the distributions of residuals at different data points may not necessarily follow an expected pattern, even if random errors are identically distributed. For example, in the conditional residuals, the random effect estimate $\hat{\boldsymbol{b}}_i$ heavily depends on the normality assumption and is also influenced by the assumed covariance structure $\hat{V}_i$.
- A first adjustment considered in the literature is to standardize/studentize the raw residuals
- Standardized residuals are particularly important in the longitudinal data with distinctive outliers.

Scaled residuals

- A more refined type of standardized residuals is the scaled residuals. If $\boldsymbol{V}_i(\boldsymbol{\alpha})$ is the true error structure, then

$$\text{var}\,(\boldsymbol{e}_i) = \boldsymbol{V}_i \text{ and } \text{var}\,(\widehat{\boldsymbol{e}}_i) \approx \boldsymbol{V}_i(\widehat{\boldsymbol{\alpha}})$$

  However, residuals should behave as pure measurement error, and therefore, they should reflect variability that is not explained by the specified random effects or the covariance parameters

- Thus, it is necessary to eliminate all sources of correlation by appropriate scaling to check residuals. The classical approach in this regard is to apply the Cholesky decomposition for the generation of transformed residuals that have constant variance and zero correlation.

Scaled residuals

- We use Cholesky decomposition to obtain residuals that have constant variance and zero correlation. Let $\boldsymbol{L}_i$ be a lower-triangular matrix such that

$$\hat{\boldsymbol{V}}_i = \boldsymbol{L}_i \boldsymbol{L}_i^{\mathrm{T}}$$

- Based on this decomposition we can form

$$\widehat{e}_i^{\star} = \boldsymbol{L}_i^{-1} \widehat{\boldsymbol{e}}_i = \boldsymbol{L}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{x}_i \widehat{\boldsymbol{\beta}} \right)$$

  so that $\mathrm{var}\left(\widehat{\boldsymbol{e}}_i^{\star}\right) \approx \mathbf{I}_{n_i}$ (unit variance and zero correlation).

- This is equivalent to work with the model

$$Y_i^{\star} = \boldsymbol{x}_i^{\star} \boldsymbol{\beta} + \boldsymbol{e}_i^{\star}$$

  where $\boldsymbol{Y}_i^{\star} = \boldsymbol{L}_i^{-1} \boldsymbol{Y}_i$, $\boldsymbol{x}_i^{\star} = \boldsymbol{L}_i^{-1} \boldsymbol{x}_i$, and $\boldsymbol{e}_i^{\star} = \boldsymbol{L}_i^{-1} \boldsymbol{e}_i$

- Plots of $\widehat{e^{\star}_{ij}}$ against $x^{\star}_{ijl}$ should not show systematic patterns if the assumed linear form is correct.

- In contrast, if this scatter-plot displays a systematic trend, residuals remain correlated even with the specified random effects and/or a selected residual covariance structure

- Plots of $\widehat{e}^{\star}_{ij}$ against $x^{\star}_{ijl}$ should not show systematic patterns if the assumed linear form is correct.

- In contrast, if this scatter-plot displays a systematic trend, residuals remain correlated even with the specified random effects and/or a selected residual covariance structure
- QQ plots of $\widehat{e}^{\star}_{ij}$ or $\boldsymbol{e}_{ij}$ can be used to assess normality
- The correctness of the mean-variance relationship can be assessed by plotting $\widehat{e}^{\star 2}_{ij}$ (or $\left|\widehat{e}^{\star}_{ij}\right|$) against fitted values $\widehat{\mu}^{\star}_{ij} = \boldsymbol{x}^{\star}_{ij}\widehat{\boldsymbol{\beta}}$. Any systematic (non-horizontal) trends suggest problems
- For the LMM with $\epsilon_i|\sigma^2_{\epsilon} \sim_{iid} \mathbf{N}_{n_i}\left(\mathbf{0}, \sigma^2_{\epsilon}\mathbf{I}\right)$, one can consider standardized stage one (conditional) residuals $\widehat{\epsilon}^{\star}_{ij} = \widehat{\epsilon}_{ij}/\widehat{\sigma}_{\epsilon}$
- The constant variance assumption may be examined via a plot of $\widehat{\epsilon}^{\star 2}_{ij}$ (or $\left|\widehat{\epsilon}^{\star}_{ij}\right|$) versus $\widehat{\mu}_{ij} = \boldsymbol{x}_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{z}_{ij}\widehat{\boldsymbol{b}}_i$

Mahalanobis distance for detecting outlining individuals

- For each individual we can calculate a summary measure of multivariate distance between their observed and fitted responses, based on the Mahalanobis distance,

$$d_i = \boldsymbol{e}_i^{\star\prime}\boldsymbol{e}_i^{\star}$$

- If the model is correctly specified, the distances d_i have an approximate chi-squared distribution with degrees of freedom (df) equal to the dimension of $\boldsymbol{e}_i^*$ (i.e., df $= n_i$, the number of repeated measurements on the i^{th} subject):

$$d_i \sim \chi^2_{n_i}$$

- Outlying individuals will have distances d_i, that have small associated p -values. The p -values
- Word of caution: use a Bonferroni correction or a FDR criterion (distribution of extremes is not χ^2.

Diagnostics on serial correlation

- Recall the model

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i$$

with $\boldsymbol{b}_i|\boldsymbol{D} \sim_{iid} \mathbf{N}_{q+1}(\mathbf{0}, \boldsymbol{D})$ and $\boldsymbol{\epsilon}_i|\sigma^2_\epsilon \sim_{iid} \mathbf{N}_{n_i}\left(\mathbf{0}, \sigma^2_\epsilon\mathbf{I}\right)$ representing random effects and measurement error and δ_{ij} being zero-mean normal error terms with serial dependence in time.

- This leads to the variance-covariance for responses on unit i :

$$\text{var}(\boldsymbol{Y}_i) = \boldsymbol{V}_i = \boldsymbol{z}_i D(\boldsymbol{\alpha})\boldsymbol{z}_i' + \sigma^2_\delta \boldsymbol{R}_i + \sigma^2_\epsilon \mathbf{I}_{n_i}$$

- As summarized by Diggle (1988) and Diggle et al. (2002), an appropriate V_i should at least accommodate three different sources of random variations. First, average responses usually vary randomly between subjects, with some subjects being intrinsically high and some being low. Second, a subject's observed measurement profile may be a response to time-varying stochastic processes. Third, as the individual measurements involve some kind of subsampling within subjects, he measurement process adds a component of variation to the data.

Autocorrelation function

- We consider the model
$$Y_t = \mu_t + \epsilon_t$$
- We assume the error terms ϵ_t are second-order stationary, which means that $\mathrm{E}[\epsilon_t] = \mu$ is constant, and $\mathrm{cov}(\epsilon_t, \epsilon_{t+d}) = C(d)$, where $d \geq 0$, that is, the covariance only depends on the temporal spacing between the variables.

⇨ Variance is constant and equal to $\sigma_\epsilon + \sigma_\delta = C(0)$.

⇨ Autocorrelation function (ACF) is defined, for time points $d \geq 0$ apart, as
$$\rho(d) = \frac{\mathrm{cov}(\epsilon_t, \epsilon_{t+d})}{\sqrt{\mathrm{var}(\epsilon_t)\,\mathrm{var}(\epsilon_{t+d})}} = \frac{C(d)}{C(0)}$$

- The empirical ACF is defined as
$$\widehat{\rho}(d) = \frac{\widehat{C}(d)}{\widehat{C}(0)} = \frac{\sum_{t=1}^{n-d} \widehat{\epsilon}_t \widehat{\epsilon}_{t+d}/(n-d)}{\sum_{t=1}^{n} \widehat{\epsilon}_t^2/n}$$

for $d = 0, 1, \ldots, n-1$.

Autocorrelation function

- A correlogram plots $\widehat{\rho}(d)$ versus d for $d = 0, 1, 2, \ldots, n-1$. If the residuals are a white noise process (i.e., uncorrelated), then asymptotically

$$\sqrt{n}\,\widehat{\rho}(d) \rightarrow_d \mathrm{N}(0,1)$$

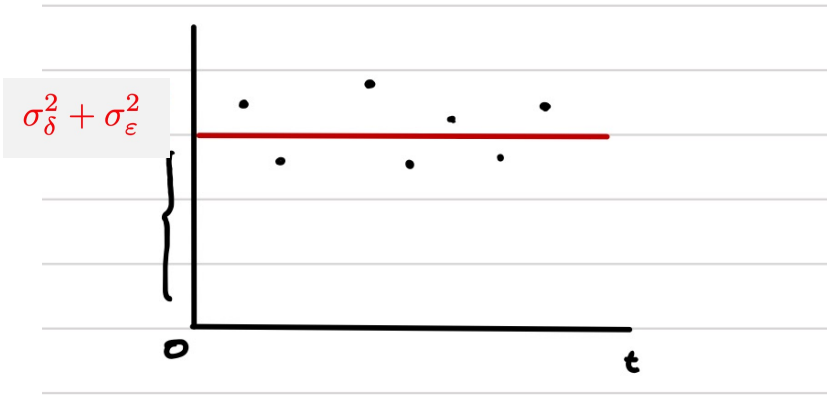
for $d = 1, 2, \ldots,$ to give, for example, 95% confidence bands of $\pm 1.96/\sqrt{n}$

Semivariogram

- A tool for spatial EDA introduced by Matheron (1971) - and Diggle et al. (2002, moved to LDA
- Basic Idea: Identify the different sources of variability in the data.
- For example, let's consider the marginal variance in a random intercept model

$$Var(Y_i) = \sigma_b^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i} + \sigma_\delta^2 R_i + \sigma_\varepsilon^2 I_{n_i}$$

If the model is correctly specified and there is no residual serial correlation, the residuals of the fit should represent only the measurement component



Semi-variogram

The semi-variogram of the residuals ϵ_t, is defined as

$$\gamma(d) = \frac{1}{2}\operatorname{var}(\epsilon_t - \epsilon_{t+d}) = \frac{1}{2}\mathrm{E}\left[(\epsilon_t - \epsilon_{t+d})^2\right]$$

for $d \geq 0$. The semi-variogram exists under weak stationarity. Then, assuming zero-mean error terms,

$$\begin{aligned}\gamma(d) &= \frac{1}{2}\operatorname{var}(\epsilon_t) + \frac{1}{2}\operatorname{var}(\epsilon_{t+d}) - \operatorname{cov}(\epsilon_t, \epsilon_{t+d}) \\ &= C(0) - C(d) \\ &= C(0)[1 - \rho(d)]\end{aligned}$$

That is, the semi-variogram increases with vanishing correlation (which is typically the case for d large). $C(0)$ is the variance of the residuals.

Sample Semi-variogram

The sample semi-variogram uses the empirical halved squared differences between pairs of residuals

$$v_{ll'} = \frac{1}{2}\left(\widehat{\epsilon}_l - \widehat{\epsilon}_{l'}\right)^2$$

along with the spacings $d_{ll'} = |t_l - t_{l'}|$ for $l = 1, \ldots, n$ and $l < l' = 1, \ldots, n$

In practice, it may be more difficult to interpret. Marginally, with normal data, the distribution of each $v_{ll'}$ is is $C(0)\chi^2_1$, hence large variability. Due to dependence, a single outlying point can influence the plot at different time lags.

Semi-variogram for LMMs

- We consider a LMM with the usual assumptions:

$$\boldsymbol{y}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

- Based on the fit, we can obtain the semi-variogram for the i-th individual

$$\gamma_i\left(d_{ijk}\right) = \frac{1}{2}\mathrm{E}\left[\left(\epsilon_{ij} - \epsilon_{ik}\right)^2\right]$$

and sample semi-variogram

$$v_{ijk} = \frac{1}{2}\left(\widehat{\epsilon}_{ij} - \widehat{\epsilon}_{ik}\right)^2$$

If no serial dependence is present, the smoother should be roughly horizontal. In other words, if the random intercept model is correctly assumed, the errors should follow a constant pattern, and, correspondingly, the empirical semi-variogram should be scattered randomly rather than systematically.

Semi-variogram for LMMs

Suppose that a random intercept model with serial correlation is the truth,

$$Y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i + \delta_{ij} + \epsilon_{ij}$$

but we fit a model without the autocorrelated terms

$$Y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i + \epsilon'_{ij}$$

Then, the stage one residuals become:

$$\epsilon'_{ij} = Y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\beta} - \boldsymbol{z}_{ij}\boldsymbol{b}_i = \delta_{ij} + \epsilon_{ij}$$

For differences in residuals on the same individual,

$$\epsilon'_{ij} - \epsilon'_{ik} = \delta_{ij} + \epsilon_{ij} - \delta_{ik} - \epsilon_{ik} = (\delta_{ij} - \delta_{ik}) + (\epsilon_{ij} - \epsilon_{ik})$$

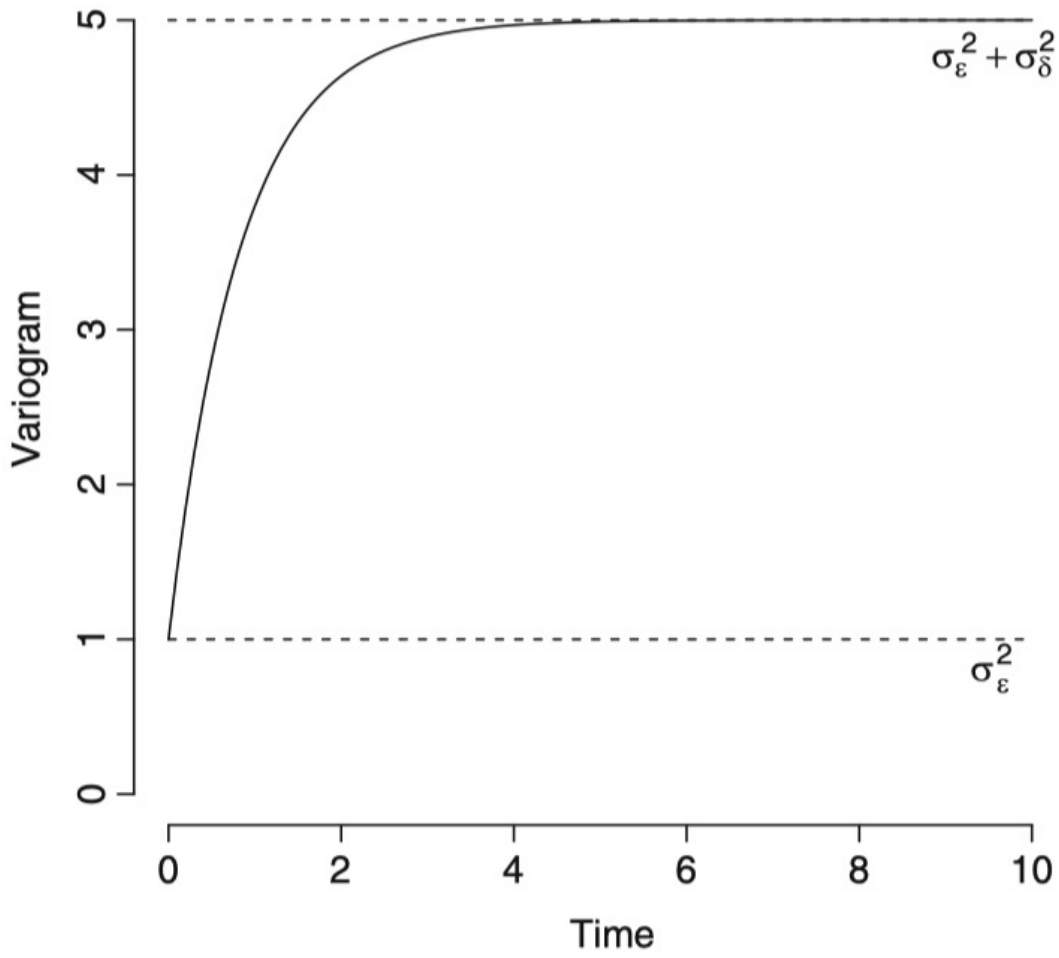
So the semi-variogram

$$\begin{aligned}\gamma_i(d_{ijk}) &= \frac{1}{2}\mathrm{E}\left[\left(\epsilon'_{ij} - \epsilon'_{ik}\right)^2\right] = \frac{1}{2}\mathrm{E}\left[\left(\delta_{ij} - \delta_{ik}\right)^2 + \left(\epsilon_{ij} - \epsilon_{ik}\right)^2\right] \\ &= \sigma^2_\delta\left[1 - \rho(d_{ijk})\right] + \sigma^2_\epsilon\end{aligned}$$

As $d_{ijk} \to 0, \gamma_i(d_{ijk}) \to \sigma^2_\epsilon$ (nugget); The rate at which asymptote $\sigma^2_\delta + \sigma^2_\epsilon$ is reached as $d_{ijk} \to \infty$ is determined by ρ (range and silll).

Theoretical Semivariogram

Fig. 8.4 Theoretical (semi-)variogram corresponding to (8.53) with $\sigma_\epsilon^2 = 1$, $\sigma_\delta^2 = 4$ and $\rho = 0.3$



Semi-variogram for population residuals

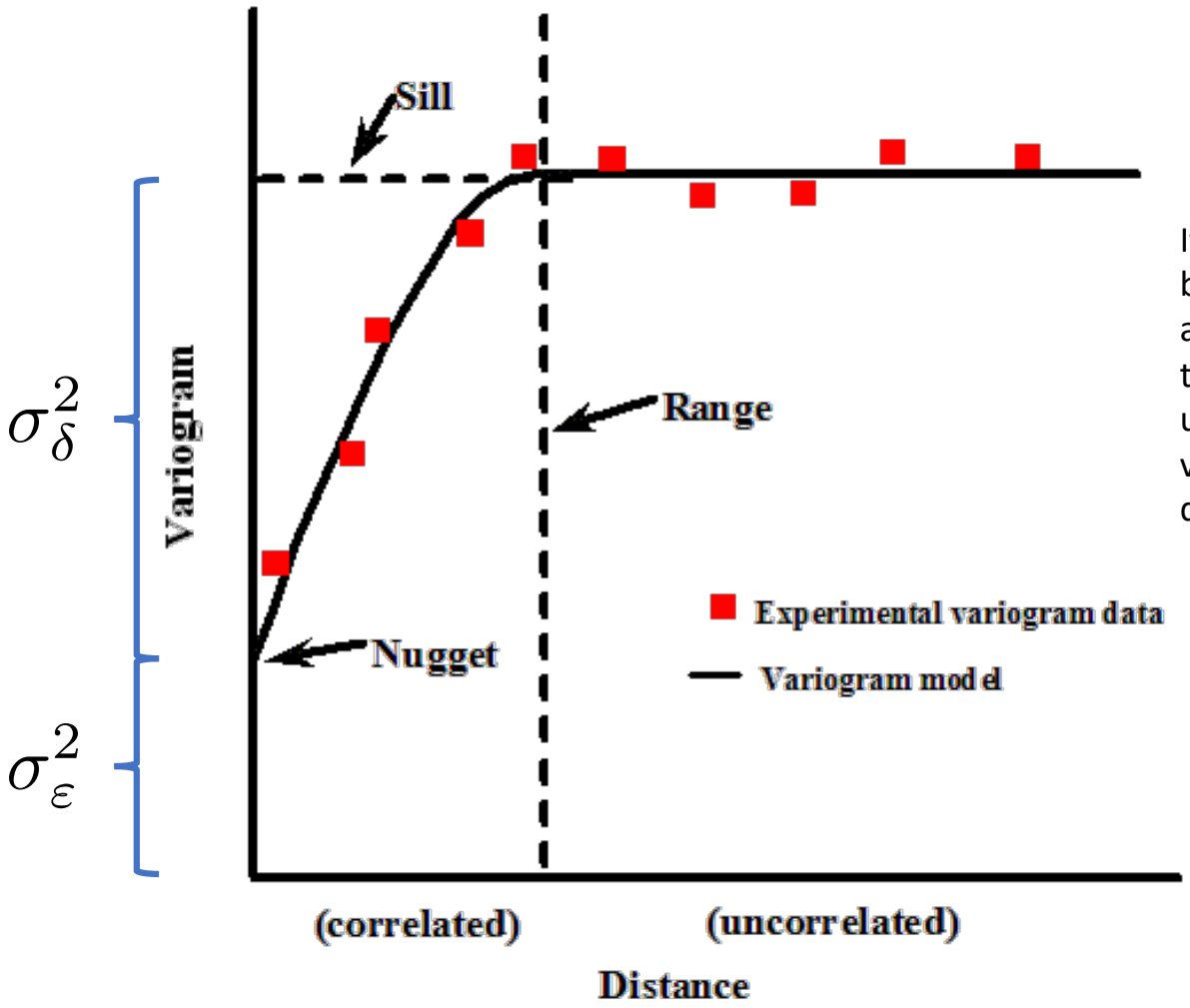
- The population (marginal) residuals under this model are

$$e_{ij} = Y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\beta} = b_i + \delta_{ij} + \epsilon_{ij}$$

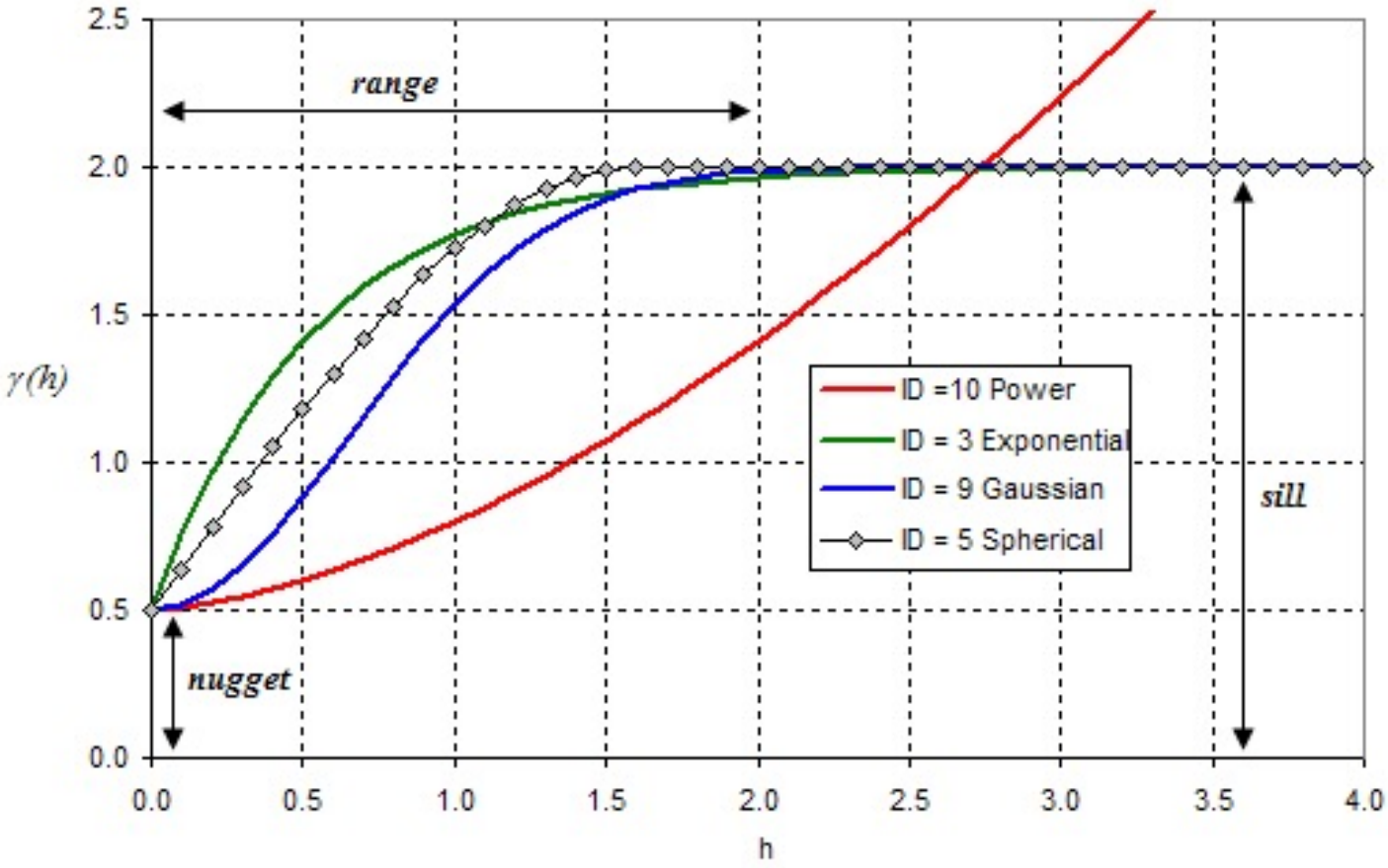
which cancel out when we consider individual differences

$$\begin{aligned} e_{ij} - e_{ik} &= b_i + \delta_{ij} + \epsilon_{ij} - b_i - \delta_{ik} - \epsilon_{ik} \\ &= (\delta_{ij} - \delta_{ik}) + (\epsilon_{ij} - \epsilon_{ik}) \end{aligned}$$

Since b_i is constant within individuals, the variance σ_b^2 is not represented.



If there is still a gap between the variogram and the sill, that means that there are still unexplained sources of variance/correlation in the data



http://xongrid.sourceforge.net/GpVariogram.htm

An exponential correlation model $e^{-\phi\, d}$ and semi-variogram $\sigma_\delta(1 - e^{-\phi\, d}) + \sigma_\varepsilon$ and the time-series realizations for different values of the parameter ϕ

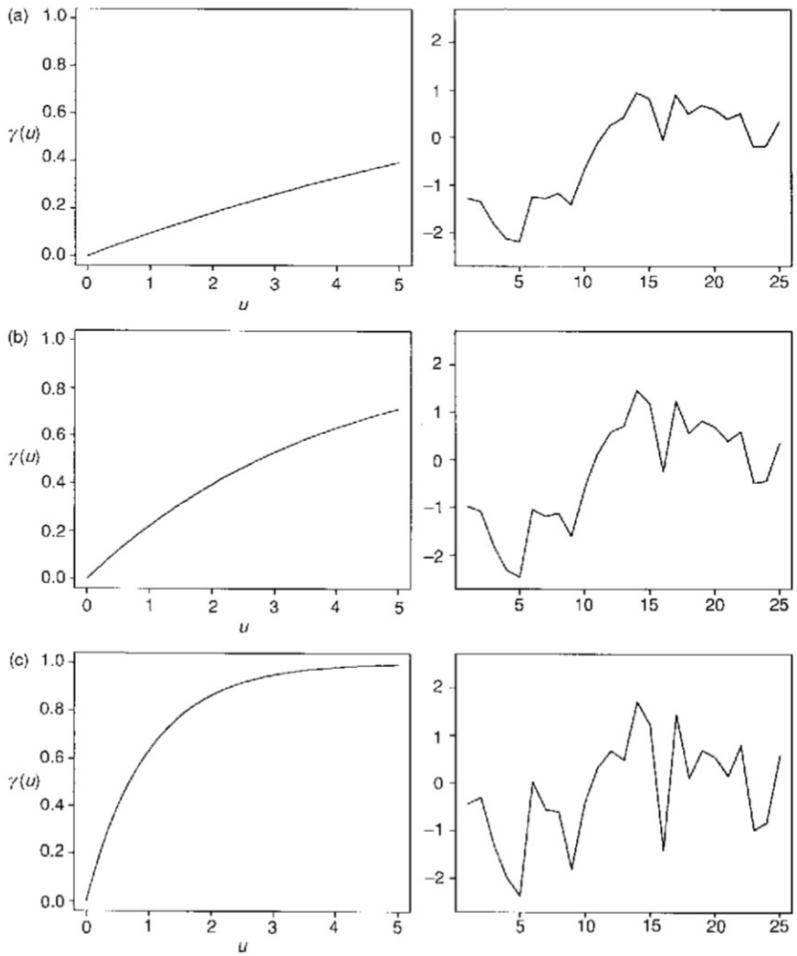
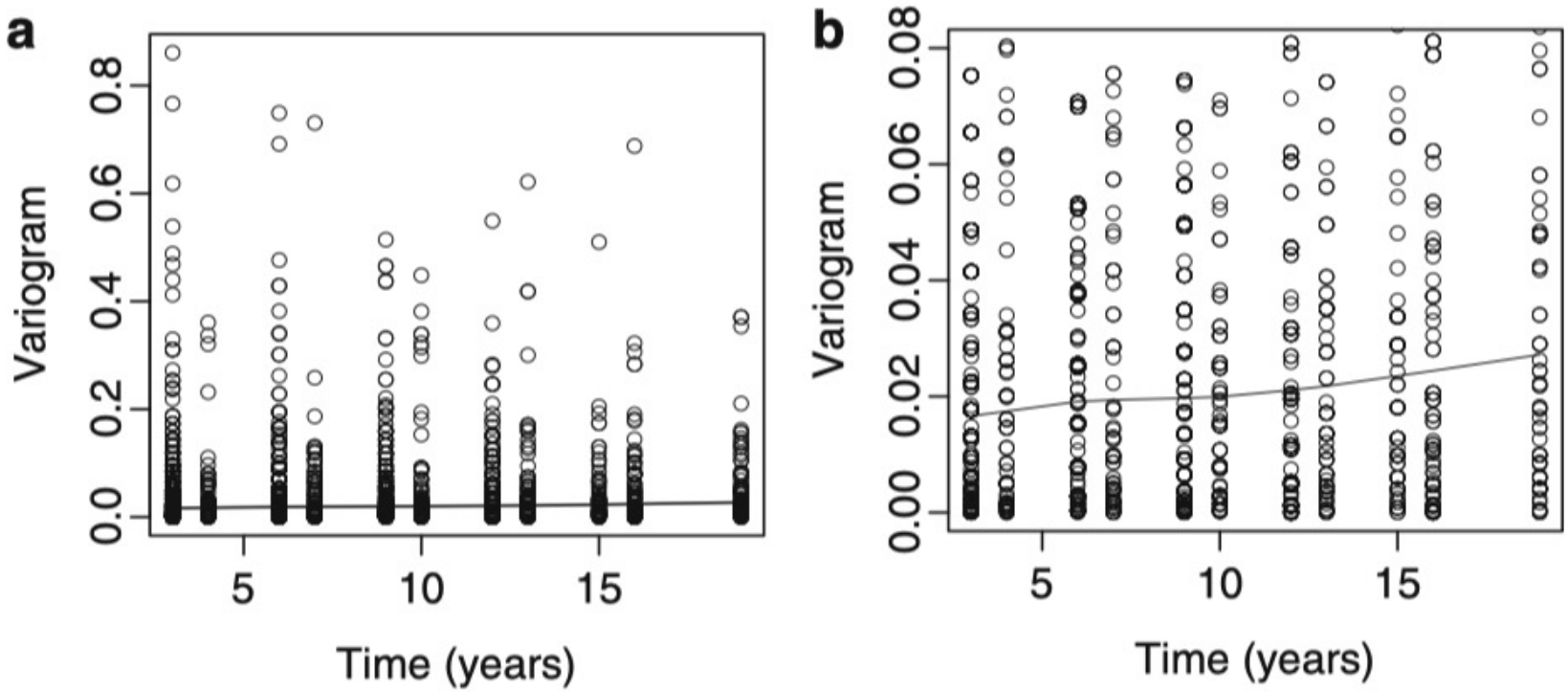


Fig. 5.1. Variograms (left-hand panels) and simulated realizations (right-hand panels) for the exponential correlation model: (a) $\phi = 0.1$; (b) $\phi = 0.25$; (c) $\phi = 1.0$.

Challenges



For the FEV1 data: (a) the (semi)-variogram of stage one residuals, (b) on a truncated semi-variogram scale show that interpretability may be difficult, depending on data

- However, things are (alto theoretically) a bit more complicated when we consider more complicated LMMs, since in that case the marginal variance varies with time: since $b_{i0}\,|D_0 \sim \mathrm{N}\,(0, D_0)\,, \quad b_{i1}|\,D_1 \sim \mathrm{N}\,(0, D_1)$ we get that the marginal variance is

$$\mathrm{var}\,(Y_{ij}) = \sigma^2_\epsilon + D_0 + D_1 t^2_{ij},$$

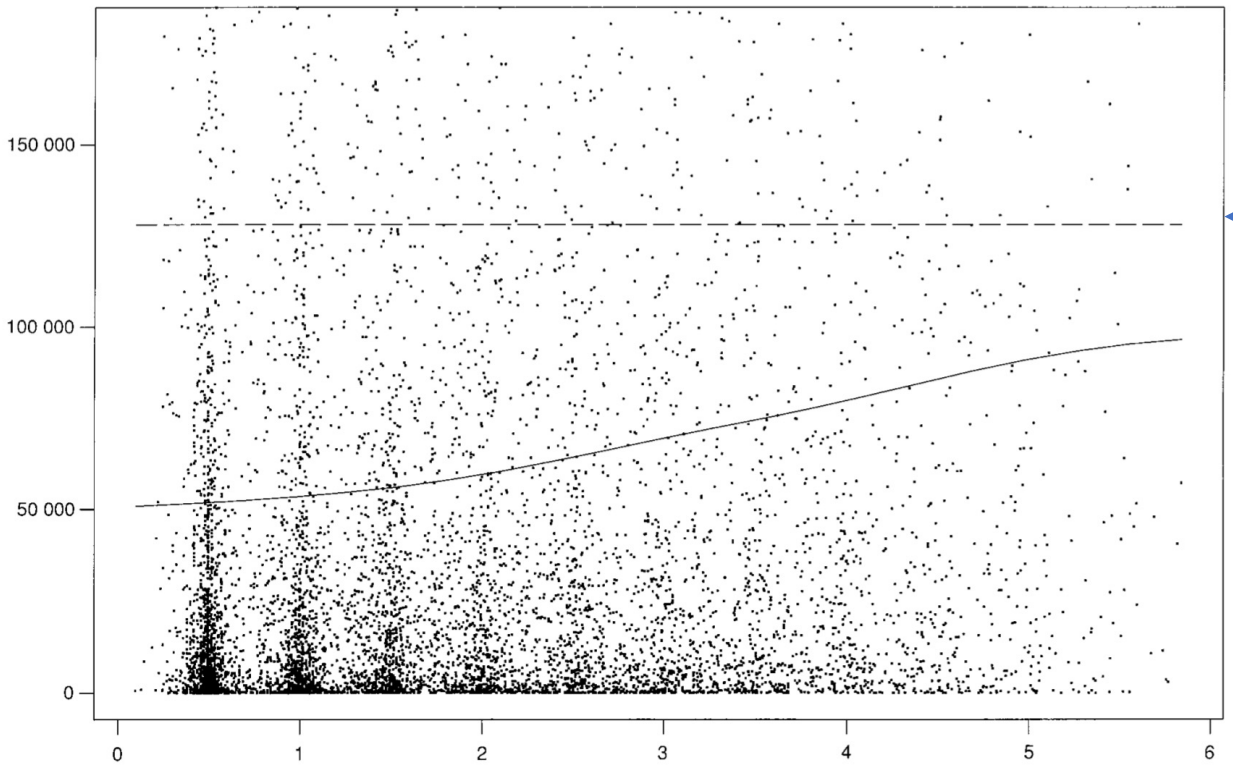
  which is non-stationary

- With the inclusion of the random coefficients, the semivariogram based on the random intercept model is extended to the random coefficient perspective. As the covariance structure is dominated by its first component $Z_i D Z_i'$, it is necessary to remove all variability explained by the random effects b_i before proceeding with the serial correlation check.

- The scaled residuals can be used for this removal, as they are independent of any distributional assumptions on $\boldsymbol{b}_i$ and thus their does not require an estimate of the random-effects covariance matrix $\boldsymbol{D}$(Verbeke et al., 1998).
- Since the scaled residuals $\hat{e}^{\star}_{ij}$ unit variance and zero correlation, the semivariogram based on the linear random coefficient model can be written as

$$\frac{1}{2}E\left[(\hat{e}^{\star}_{ij} - \hat{e}^{\star}_{ik})^2\right] = \frac{1}{2}V(\hat{e}^{\star}_{ij}) + \frac{1}{2}V(\hat{e}^{\star}_{ik}) - corr(\hat{e}^{\star}_{ij}, \hat{e}^{\star}_{ik}) = 1$$

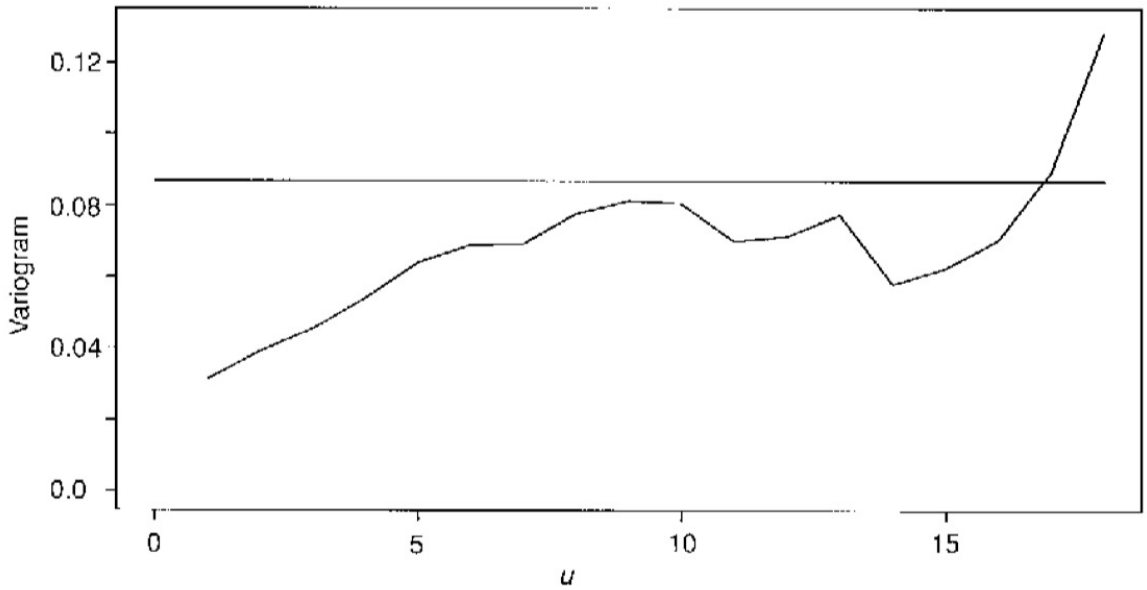
If a random coefficient linear model is correctly specified, the plot of the semivariogram of the transformed residuals against time should be scattered randomly around unity without displaying any systematic time trend. With this property, the semivariogram can be applied to check whether the specified random effects explain all serial correlation between repeated measurements using the scaled marginal (not the conditional) residuals of the LMM. With this diagnostics it is not recommended to use the conditional residuals, defined as $Y_i - X_i'\hat{\beta} - Z_i\hat{b}$, because the BLUP $\hat{b}_i$ from the empirical Bayes approximation depends heavily on the normality assumption and is also influenced by the assumed covariance structure $\hat{V}_i$.



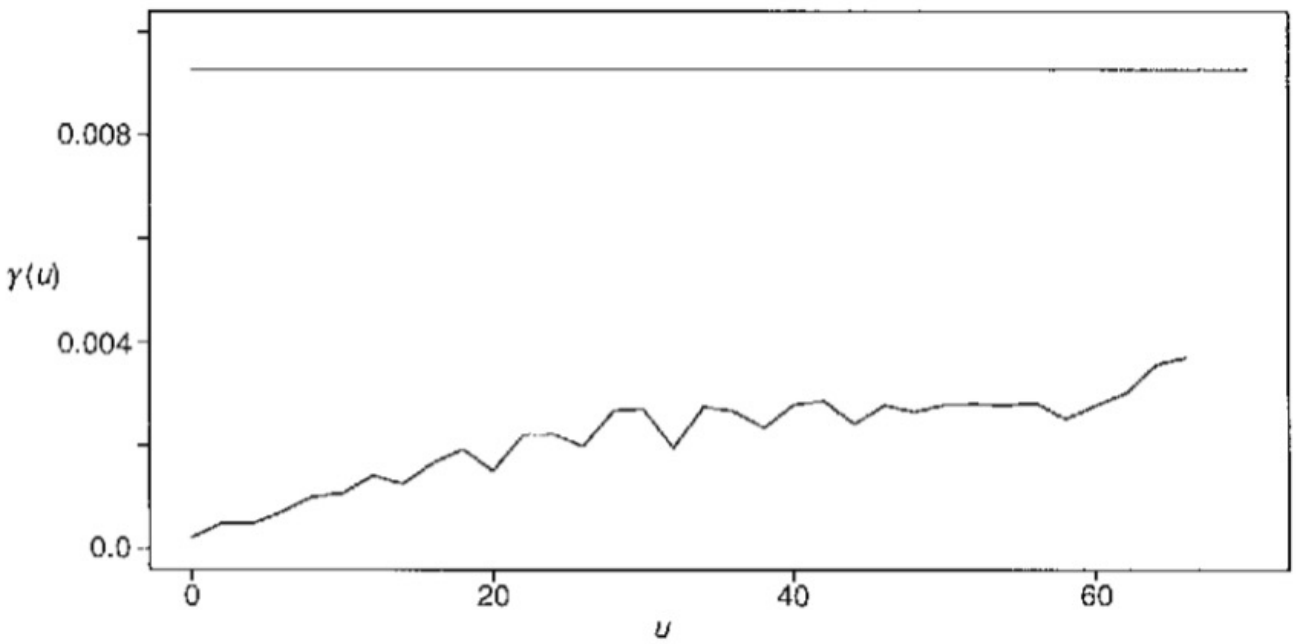
Variogram based estimate of the sample variance.

If using scaled residuals, this line should be near 1.

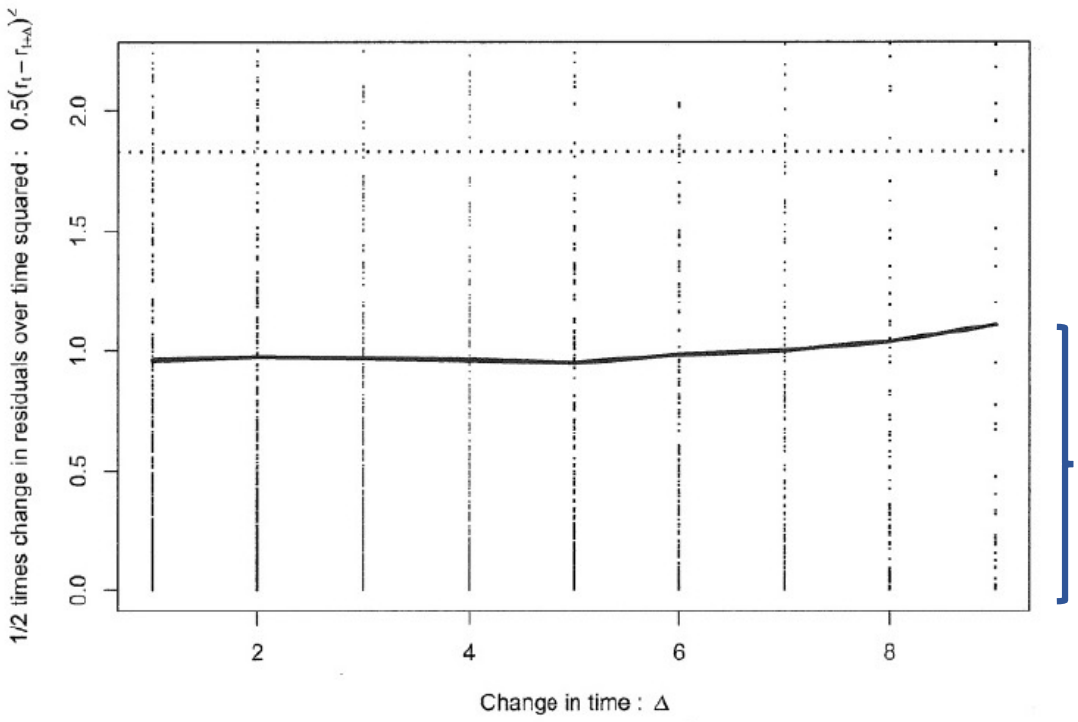
The variogram smoothly increases with lag corresponding to decreasing correlation as observations are separated in time, but appears almost to have levelled out by lag u = 6. The variogram-based estimate of the process variance is substantially larger than the value of the sample variogram at lag u = 6. This suggests that either the autocorrelation has not decayed to zero within the range of the data or that positive correlation remains at arbitrarily large time separations. In LMMs, and when using raw residuals, this could also be explained by the presence of a random effect component (e.g. intercept) that characterizes between-individual variability.



The variogram rises steadily with increasing u, but levels off within the range spanned by the data. The last increase is spurious and due to an estimation based on just a few data only



The measurement error variance seems to be very small, whereas the partial sill is quite large.
It looks like there is still residual correlation at lag 60
The shape of the variogram function suggests either an exponential or gaussian correlation structure, possibly gaussian based on the way the variogram approaches zero



Variogram computed from the raw residuals, not the scaled residuals

Measurement error component

There is a constant positive association between data which is not accounted for and does not decay even at large lags

Model Selection of GEE

- There are several reasons why model selection of GEE models is important and necessary:

  1 GEE has gained increasing attention in biomedical studies which may include a large group of predictors. Therefore, variable selection is necessary for determining which are included in the final regression model by identifying significant predictors

  2 We have seen that one feature of GEE is that the consistency of parameter estimates can still hold even when the "working" correlation structure is mis-specified. But, correctly specifying the "working" correlation structure can definitely enhance the efficiency of the parameter estimates in particular when the sample size is not large enough

  3. The variance function $v(\mu)$ is another potential factor affecting the goodness-of-fit of GEE. Correctly specified variance function can assist in the selection of covariates and an appropriate correlation structure

Quasi-likelihood under the independence model criterion

- AIC is a well-known criterion for likelihood-based model selection. However, we cannot apply a criterion such as AIC to the GEE approach, since the GEE is not likelihood-based.
- Pan(2001) proposed a criterion based on quasi-likelihood, named QIC, to select the proper mean model or the working correlation structure.
- In the linear case, the quasi-likelihood function on individual (cluster) i and observation t evaluated at the regression parameters $\boldsymbol{\beta}$ is given by $Q\left(\boldsymbol{\beta}, \phi; Y_{it}, \mathbf{x}_{it}\right) = Q_{it}/\phi = -\frac{1}{2}\left(y_{it} - \mu_{it}\right)^2$

- The basic idea is to calculate the expected Kullback-Leibler discrepancy using the quasi-likelihood under the independence "working" correlation assumption due to the lack of a general and tractable quasi-likelihood for the correlated data under any other complex "working" correlation structures.
- The QIC is expressed as

$$\text{QIC}(\mathbf{R}) = -2\sum_{i=1}^{K}\sum_{t=1}^{n} Q\left(\boldsymbol{\beta}, \phi; Y_{it}, \mathbf{x}_{it}\right) + 2\,\text{tr}\left\{\boldsymbol{\Omega}\mathbf{V}_{\mathbf{r}}(\mathbf{R})\right\}$$

where $\boldsymbol{R}$ is the hypothesized working correlation structure, $\mathbf{V}_{\mathbf{r}}(\mathbf{R})$ indicates the sandwiched variance-covariance estimator of $Cov(\hat{\beta})$ under the hypothesized working correlation structure, and

$\Omega = \sum_{i=1}^{K} \mathbf{D}_i^{\top}\mathbf{A}_i^{-1}\mathbf{D}_i$ with $\mathbf{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}$ and $\mathbf{A}_i = \text{diag}\left(v_{it}\right)$
$(\mathbf{V}_i = \phi\mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{\frac{1}{2}})$

- Note that QIC(R) do not perform well in distinguishing the independence and exchangeable "working" correlation structures

- The attractive property of the QIC criterion is that it allows the selection of the covariates and "working" correlation structure simultaneously. However, Hin and Wang (2009) noted that QIC is undermined by the first term, since it is theoretically independent of the correlation structure but depends on the mean which has been estimated under the hypothesized independent "working" correlation structure $\boldsymbol{R}$.

Correlation Information Criterion (CIC)

- To better select a "working" correlation structure, Hin and Wang (2009) propose CIC as a modification of QIC to improve its performance,

$$\text{CIC}(\mathbf{R}) = \text{tr}\left\{\boldsymbol{\Omega}\mathbf{V}_{\text{r}}(\mathbf{R})\right\}$$

- CIC is constructed using only the second term that represents the penalty of QIC. The first term in QIC denotes the sum of quasi-likelihoods for all observations under the assumption that the subjects and time points are independent.
- In their work, CIC was shown to outperform QIC
- One limitation of this criterion is that it cannot penalize the over-parameterization; thus the performance is not great in comparison when you are dealing with two correlation structures having quite different numbers of correlation parameters.