MATH 60604A
Statistical modelling
§ 5f - Other covariance models

Léo Belzile

HEC Montréal
Department of Decision Sciences

Heterogeneous autoregressive structure

- With longitudinal data, it often happens that the variance of the observations is a function of the measurement time.
- Several covariance structures in `proc mixed` have a "heterogeneous" version, i.e., where the variances can be different for different time measures.
- The heterogeneous autoregressive model, $\mathsf{ARH}(1)$, has the same correlation structure as $\mathsf{AR}(1)$. Its covariance matrix is

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 & \sigma_1\sigma_5\rho^4 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 & \sigma_2\sigma_5\rho^3 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho & \sigma_3\sigma_5\rho^2 \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 & \sigma_4\sigma_5\rho \\ \sigma_5\sigma_1\rho^4 & \sigma_5\sigma_2\rho^3 & \sigma_5\sigma_3\rho^2 & \sigma_5\sigma_2\rho & \sigma_5^2 \end{pmatrix}.$$

- Instead of assuming a common variance σ^2 for all time measures, the $\mathsf{ARH}(1)$ model allows for different variance σ_j^2 at time j.

Syntax for adjusting the heterogeneous ARH(1) model

SAS code to fit heterogeneous ARH(1) model

```
proc mixed data=revenge method=reml;
class id tcat;
model revenge = sex age vc wom t / solution;
repeated tcat / subject=id type=arh(1) r=1 rcorr=1;
run;
```

Correlation and covariance matrices for subject 1 (ARH(1))

Estimated R Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.2937	0.1756	0.09260	0.05184	0.02271
2	0.1756	0.3937	0.2077	0.1162	0.05093
3	0.09260	0.2077	0.4109	0.2300	0.1008
4	0.05184	0.1162	0.2300	0.4829	0.2116
5	0.02271	0.05093	0.1008	0.2116	0.3477

Estimated R Correlation Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.5163	0.2666	0.1376	0.07107
2	0.5163	1.0000	0.5163	0.2666	0.1376
3	0.2666	0.5163	1.0000	0.5163	0.2666
4	0.1376	0.2666	0.5163	1.0000	0.5163
5	0.07107	0.1376	0.2666	0.5163	1.0000

- The covariance matrix shows that the variance of the observations is different for each measurement time
- The correlation matrix shows, as for the $\mathrm{AR}(1)$ structure, that the correlation between two observations decreases over time.

Covariance parameters estimates of ARH(1) model

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Var(1)	id	0.2937
Var(2)	id	0.3937
Var(3)	id	0.4109
Var(4)	id	0.4829
Var(5)	id	0.3477
ARH(1)	id	0.5163

- This model includes six parameters in the covariance structure.
- In this table, we can see estimates of the variances of the five time points.
- The estimate for ρ is $\widehat{\rho} = 0.516$, comparable to the estimate from the preceding model (0.492).

Likelihood ratio test for ARH(1)

Estimated R Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.3133	0.2165	0.07287	0.05930	0.03405
2	0.2165	0.4233	0.2258	0.1355	0.07025
3	0.07287	0.2258	0.4204	0.2371	0.1689
4	0.05930	0.1355	0.2371	0.4508	0.1444
5	0.03405	0.07025	0.1689	0.1444	0.3179

Estimated R Correlation Matrix for id 1					
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.5946	0.2008	0.1578	0.1079
2	0.5946	1.0000	0.5354	0.3103	0.1915
3	0.2008	0.5354	1.0000	0.5446	0.4621
4	0.1578	0.3103	0.5446	1.0000	0.3815
5	0.1079	0.1915	0.4621	0.3815	1.0000

- SAS outputs the results from the likelihood ratio test comparing the ARH(1) model (complete model) with a classical linear regression model with no correlation structure (reduced model).
- The hypotheses are $\mathscr{H}_0 : \rho = 0, \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$ and $\mathscr{H}_1 : \rho \neq 0$ or at least one variance differs.
- The p-value is negligible, thus we can reject $\mathscr{H}_0$ and conclude in favour of the ARH(1) correlation structure.

Other possibilities for the choice of covariance structure

- Another possibility is not specifying any structure on the covariance,

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54}^2 & \sigma_5^2 \end{pmatrix}$$

- This kind of structure can often be used to explore the covariance structure without imposing any specific restrictions. We can only use this structure when the number of observations in each group is small and the number of groups is large, because it has $n \times (n+1)/2$ parameters.
- In our example, we get 15 parameters, compared to two parameters for the compound symmetry and the $\mathrm{AR}(1)$ covariance models, and to six for the $\mathrm{ARH}(1)$ covariance model.

Unstructured covariance model

SAS code to adjust an unstructure covariance model

```
proc mixed data=revenge method=reml;
class id tcat;
model revenge = sex age vc wom t / solution;
repeated tcat / subject=id type=un r=1 rcorr=1;
run;
```

Estimated correlation and covariance matrices for subject 1

Estimated R Matrix for id 1

Row	Col1	Col2	Col3	Col4	Col5
1	0.3133	0.2165	0.07287	0.05930	0.03405
2	0.2165	0.4233	0.2258	0.1355	0.07025
3	0.07287	0.2258	0.4204	0.2371	0.1689
4	0.05930	0.1355	0.2371	0.4508	0.1444
5	0.03405	0.07025	0.1689	0.1444	0.3179

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.5946	0.2008	0.1578	0.1079
2	0.5946	1.0000	0.5354	0.3103	0.1915
3	0.2008	0.5354	1.0000	0.5446	0.4621
4	0.1578	0.3103	0.5446	1.0000	0.3815
5	0.1079	0.1915	0.4621	0.3815	1.0000

- The variances are different for each time point, and there is no specific structure for the correlations.
  1. The estimated variances are very similar for all five time points.
  2. The estimated correlations between observations seems to decrease as the time between measurements increases.
- This would suggest that the $\mathrm{AR}(1)$ structure is preferable to the compound symmetry structure.

Estimates of the unstructured covariance parameters

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
UN(1,1)	id	0.3133
UN(2,1)	id	0.2165
UN(2,2)	id	0.4233
UN(3,1)	id	0.07287
UN(3,2)	id	0.2258
UN(3,3)	id	0.4204
UN(4,1)	id	0.05930
UN(4,2)	id	0.1355
UN(4,3)	id	0.2371
UN(4,4)	id	0.4508
UN(5,1)	id	0.03405
UN(5,2)	id	0.07025
UN(5,3)	id	0.1689
UN(5,4)	id	0.1444
UN(5,5)	id	0.3179

Information criteria and likelihood ratio test

Fit Statistics	
-2 Res Log Likelihood	659.3
AIC (Smaller is Better)	689.3
AICC (Smaller is Better)	690.6
BIC (Smaller is Better)	725.0

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
14	117.34	<.0001

- As before, the AIC and BIC can be used to compare models with different correlation structures
- The likelihood ratio tests the hypothesis that the model with independent homoscedastic observations for the group, with σ^2 for all diagonal elements and zero elsewhere is adequate. The hypothesis is soundly rejected.