MATH 60604A
Statistical modelling
§ 5g - Selection of the covariance structure

HEC Montréal
Department of Decision Sciences

Review of the covariance models covered

- We have seen how to fit five kinds of covariance structures on the errors in the regression model:
  - The CS (exchangeable) structure: each pair of observations has the same correlation
  - The AR(1) structure: the correlation between two observations decreases the further apart they are in time.
  - The ARH(1) structure: same thing as AR(1), but allows a different variance at each measurement time.
  - The unstructured model, which allows a different covariance for each pair of observations in time.

Covariance structure selection using information criteria

Model	$-2\ell_{\text{reml}}$	AIC	BIC
Ordinary Regression	776.7	778.7	782.6
Compound Symmetry	709.4	713.4	718.2
AR(1)	681.8	685.8	690.5
ARH(1)	675.3	687.3	701.6
Unstructured	659.3	689.3	725.0

- Both AIC and BIC criteria point towards the AR(1) model as the "best" model.
- This is a parsimonious model (only two parameters on the covariance structure) which seems to do a good job in accounting for within-subject correlation.
- What's even more reassuring is that, no matter what covariance structure we used, the conclusions for the predictor variables remain unchanged.

A word of caution regarding tests for REML estimation

- There is a small technical detail here. If we want to compare models using the AIC or BIC and the REML estimation method is used, we must compare models with the same fixed effect variables, that is, the same predictor variables.
- The AIC and BIC coming from models with different fixed effects, and which were estimated using REML, are not comparable. They can, however, be compared using the maximum likelihood estimator.

Covariance as a nuisance parameter

- Most of the time, research questions concern the β parameters modelling the mean of the model.
- For correlated data, we now know that the covariance must be adequately modelled in order for model inference to be valid.
- Choosing the covariance structure based on criteria such as AIC and BIC is reasonable. We can, however, perform formal hypothesis tests on the covariance structure.

Reminder on restricted likelihood ratio test

- The test compares the restricted log-likelihood of the "complete" model (under $\mathscr{H}_1$) with that of the "reduced" **nested** model (under $\mathscr{H}_0$).
- The null hypothesis is that the reduced model is an adequate simplification of the complete model.
- The likelihood ratio test statistic is

$$D = 2\{\ell_{\mathsf{reml}}(\widehat{\boldsymbol{\theta}}) - \ell_{\mathsf{reml}}(\widehat{\boldsymbol{\theta}}_0)\}$$

- Under $\mathscr{H}_0$, $D \stackrel{\cdot}{\sim} \chi^2_k$, where the degrees of freedom k is the difference in the number of parameters in the two models.
- We calculate the p-value for this test by using the χ^2_k distribution.

Likelihood ratio test for covariance structures

- It's possible to test hypotheses involving complex models using the likelihood ratio test.
- For example, we can test if it's necessary to have different variances in the model for an AR(1).
- In this case, we would want to test if the AR(1) model is adequate or if the ARH(1) model is necessary.
- The hypotheses are $\mathscr{H}_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$ against the alternative that at least two variances are different.
- The complete model is thus the model with ARH(1) structure ($\mathscr{H}_1$) and the reduced model is that with the AR(1) structure ($\mathscr{H}_0$).

Likelihood ratio test to compare AR(1) and ARH(1)

- Based on the tables reported previously, the difference in restricted log-likelihood $-2\ell_{\text{reml}}$ for these two models is $681.8 - 675.3 = 6.5$.
- There are **four** additional parameters in the complete model. We compare to a χ_4^2, whose 95% quantile is 9.48.

> ### SAS code to compute p-value using the χ_4^2 null distribution
>
> ```
> data pval;
> pval=1-CDF('CHISQ',6.5,4);
> run;
> proc print data=pval;
> run;
> ```

- We obtain a p-value of 0.165. Thus, we fail to reject $\mathcal{H}_0$ and conclude that the AR(1) model seems adequate.

Final remark

- We used REML throughout for estimating the variance parameters (default option of `proc mixed`).
- Many models are nested so use formal likelihood ratio tests whenever possible for comparisons.
  - e.g., independence $\prec$ AR(1) $\prec$ ARH(1) $\prec$ unstructured.
- Using AIC or BIC to compare models is valid **provided** the mean model includes the **same** variables, as was the case for the models so far in this chapter.
- If we wanted to use the AIC and BIC to compare models with different variables for the mean, then we would need to use maximum likelihood, rather than REML.