HW5 (Due May 9)

1. Reproduce Tables 30, 31, 32, 33, 34 and comment.

Table 30: Unstructured covariance

Marginal variance covariance matrix

[0] [4] [6] [8] [12]

[0] 9.6683 10.175 8.9741 9.8125 9.4071

[4] 10.1750 12.550 11.0910 12.5800 11.9280

[6] 8.9741 11.091 10.6420 11.6860 11.1010

[8] 9.8125 12.580 11.6860 13.9910 13.1210

[12] 9.4071 11.928 11.1010 13.1210 13.9450

Here we see that variances are not homogeneous through time

Table 31: Unstructured correlation

[0] [4] [6] [8] [12]

[0] 1.00000 0.92373 0.88473 0.84370 0.81017

[4] 0.92373 1.00000 0.95973 0.94939 0.90170

[6] 0.88473 0.95973 1.00000 0.95771 0.91127

[8] 0.84370 0.94939 0.95771 1.00000 0.93942

[12] 0.81017 0.90170 0.91127 0.93942 1.00000

Here we see that correlation tends to decrease as observations are more spaced in time.

Table 32: AR1 correlation structure

[0] [4] [6] [8] [12]

[0] 1.00000 0.94018 0.88393 0.83105 0.78133

[4] 0.94018 1.00000 0.94018 0.88393 0.83105

[6] 0.88393 0.94018 1.00000 0.94018 0.88393

[8] 0.83105 0.88393 0.94018 1.00000 0.94018

[12] 0.78133 0.83105 0.88393 0.94018 1.00000

Here we see that, although this correlation structure accounts for drop in correlation as time passes, the drop is to abrupt when compared to unstructured correlation structure

Table 33: Exponential correlation structure

[0] [4] [6] [8] [12]

[0] 1.00000 0.91694 0.87804 0.84079 0.77096

[4] 0.91694 1.00000 0.95757 0.91694 0.84079

[6] 0.87804 0.95757 1.00000 0.95757 0.87804

[8] 0.84079 0.91694 0.95757 1.00000 0.91694

[12] 0.77096 0.84079 0.87804 0.91694 1.00000

Here the drop in correlation seems even more sudden.

Table 34:

model dev aic

1 UN 597.3399 647.3399

2 AR1 621.0718 645.0718

3 Exp 618.5459 642.5459

AIC’s are different from table 34. I estimated them by hand and the ones presented above seem correct

e.g. For the UN covariance structure (with 25 parameters: 10 regression coefficients, 5 variances, and 10 covariances)

AIC-UN = 597.3399 + 2X25 = 647.3399

This is different from 627.3 presented in table 34.

Anyway, the exponential correlation structure seems to produce a more parsimonious fit to the data.

1. Compute the estimated variance and correlation coefficient in S296 and 298.

Not so sure what you mean by “computing”, here’s what R gave me:

AR1 model:

Var (from the covariance matrix of the model)

11.8670

Rho (from the model summary):

0.9401763

Exp model:

Var:

11.8750

Rho: Was not able to get this from R output. So I just estimated it like this:

Rho^4 = 0.91694

From the correlation matrix for observations separated by 4 units of time. Then, solving for Rho, I got:

Rho = 0.91694^-4 = 0.9785 ~ 0.98

I’m sure there must be a smarter way (?) to do it in R

1. Conduct the LRT tests in S300 and S301.

LRT between UN and AR1

Model df AIC BIC logLik Test L.Ratio p-value

m1 1 25 647.3399 724.6837 -298.6699

m2 2 12 645.0718 682.1968 -310.5359 1 vs 2 23.73189 0.0337

LRT between UN and Exp

Model df AIC BIC logLik Test L.Ratio p-value

m1 1 25 647.3399 724.6837 -298.6699

m3 2 12 642.5459 679.6709 -309.2729 1 vs 2 21.20597 0.069

1. If you were the data analyst for this study, how will you estimate the covariance pattern in the final model, and how will you justify? [Remark: You may be guided by Table 35 and/or other accepted guidelines/advice.]

Since we have observations on only 37 subjects I would not use the unstructured correlation structure, since a model with this structure has too many parameters (10 coefficients, 10 covariances, and 5 variances). Son many parameters could lead to unstable estimates of effect (our main interest here). I wouldn’t advice to use neither AR1 nor Exp, in spite of the last one not being significantly different from the model with the unstructured covariance pattern. Because first, it is borderline non-significant, and second (and as the AR1) it assumes constant variance, which is unrealistic, based on variance estimates from the unstructured covariance pattern model. Hence, I would use something in between, like the heterogeneous exponential covariance pattern, where variances are allowed to vary across measurement occasions. Here’s the LRT comparing the models’ fit with unstructured covariance and heterogeneous exponential covariance

Model df AIC BIC logLik Test L.Ratio p-value

m1 1 25 647.3399 724.6837 -298.6699

m4 2 16 637.3912 686.8912 -302.6956 1 vs 2 8.05127 0.529

There’s a non-significant drop of fit (p-value = 0.529) with the latter model, indicating that this covariance model provides a better-more-parsimonious fit to the data at hand.

1. Compare and critique on AIC vs. BIC, e.g., math formulations and pros and cons. [Remark: You can use textbooks or internet. Note that you will see QIC later when you will learn GEE!]

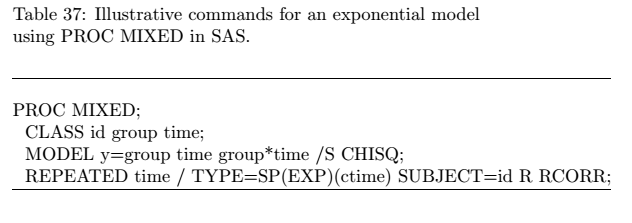
AIC = -2(logL) + 2 X number of parameters

BIC = -2(logL) + logN\*(number of parameters)

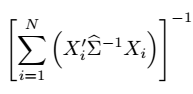
Where N\* is the number of effective subjects: N for ML estimation and N-p for REML estimation, where p is the number of regression coefficients to be estimated.

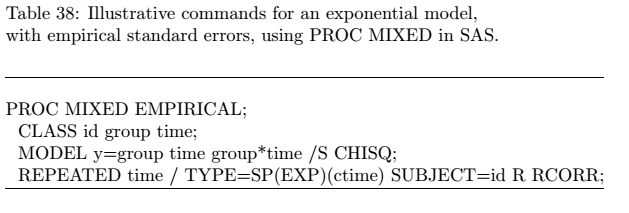
So, for N-p > 100, BIC exerts a larger penalty for each additional covariance parameter when compared to AIC. This would produce results that could be too conservative.

1. Run the codes in Tables 37 vs. 38 using the same dataset. Describe the differences of mathematical formulations and results.



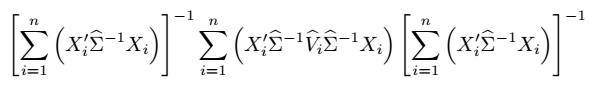
This code runs the model with an exponential covariance pattern, in which the estimator of the Cov(Yi) is , which yields the estimator of the variance of as:





This code runs a model using empirical standard errors, which, instead of using for estimating Cov (Yi), uses

This yields the following estimator of the variance of :



Now, I was not able to do it in R (without spending a whole day on it…there’s a package called sandwich. In the author’s website they resort to GEE for getting the sandwich estimator…not so sure how OK that is). Here are the results of both models in SAS:

Inferences based model SE’s

| **Type 3 Tests of Fixed Effects** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **Chi-Square** | **F Value** | **Pr > ChiSq** | **Pr > F** |
| **group** | 1 | 35 | 1.70 | 1.70 | 0.1917 | 0.2002 |
| **time** | 4 | 128 | 28.18 | 7.04 | <.0001 | <.0001 |
| **group\*time** | 4 | 128 | 3.57 | 0.89 | 0.4679 | 0.4711 |

Only time effect is significant

| **Solution for Fixed Effects** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **group** | **time** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  |  | 83.0052 | 0.7769 | 35 | 106.84 | <.0001 |
| **group** | **1** |  | -1.7099 | 1.1634 | 35 | -1.47 | 0.1506 |
| **group** | **2** |  | 0 | . | . | . | . |
| **time** |  | **0** | -1.9576 | 0.5452 | 128 | -3.59 | 0.0005 |
| **time** |  | **2** | -1.0435 | 0.4689 | 128 | -2.23 | 0.0278 |
| **time** |  | **3** | -0.3862 | 0.4196 | 128 | -0.92 | 0.3591 |
| **time** |  | **4** | -0.4593 | 0.3648 | 128 | -1.26 | 0.2102 |
| **time** |  | **6** | 0 | . | . | . | . |
| **group\*time** | **1** | **0** | 0.3498 | 0.8032 | 128 | 0.44 | 0.6639 |
| **group\*time** | **1** | **2** | 0.5606 | 0.6812 | 128 | 0.82 | 0.4121 |
| **group\*time** | **1** | **3** | 0.08879 | 0.6059 | 128 | 0.15 | 0.8837 |
| **group\*time** | **1** | **4** | 0.4234 | 0.5172 | 128 | 0.82 | 0.4145 |
| **group\*time** | **1** | **6** | 0 | . | . | . | . |
| **group\*time** | **2** | **0** | 0 | . | . | . | . |
| **group\*time** | **2** | **2** | 0 | . | . | . | . |
| **group\*time** | **2** | **3** | 0 | . | . | . | . |
| **group\*time** | **2** | **4** | 0 | . | . | . | . |
| **group\*time** | **2** | **6** | 0 | . | . | . | . |

Empirical SE’s:

| **Type 3 Tests of Fixed Effects** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Num DF** | **Den DF** | **Chi-Square** | **F Value** | **Pr > ChiSq** | **Pr > F** |
| **group** | 1 | 35 | 1.66 | 1.66 | 0.1974 | 0.2059 |
| **time** | 4 | 128 | 34.41 | 8.60 | <.0001 | <.0001 |
| **group\*time** | 4 | 128 | 3.13 | 0.78 | 0.5360 | 0.5382 |

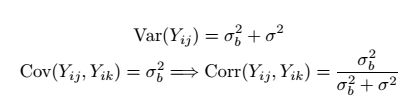
Still time is significant, but now the F statistic has decreases for both group and group-time interaction, yielding a larger p-value. On the other hand, the F statistic for the time effect has increased.

| **Solution for Fixed Effects** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **group** | **time** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** |  |  | 83.0052 | 0.6930 | 35 | 119.78 | <.0001 |
| **group** | **1** |  | -1.7099 | 1.2538 | 35 | -1.36 | 0.1813 |
| **group** | **2** |  | 0 | . | . | . | . |
| **time** |  | **0** | -1.9576 | 0.4236 | 128 | -4.62 | <.0001 |
| **time** |  | **2** | -1.0435 | 0.3597 | 128 | -2.90 | 0.0044 |
| **time** |  | **3** | -0.3862 | 0.2806 | 128 | -1.38 | 0.1711 |
| **time** |  | **4** | -0.4593 | 0.2614 | 128 | -1.76 | 0.0813 |
| **time** |  | **6** | 0 | . | . | . | . |
| **group\*time** | **1** | **0** | 0.3498 | 0.7452 | 128 | 0.47 | 0.6396 |
| **group\*time** | **1** | **2** | 0.5606 | 0.5576 | 128 | 1.01 | 0.3166 |
| **group\*time** | **1** | **3** | 0.08879 | 0.5376 | 128 | 0.17 | 0.8691 |
| **group\*time** | **1** | **4** | 0.4234 | 0.4644 | 128 | 0.91 | 0.3637 |
| **group\*time** | **1** | **6** | 0 | . | . | . | . |
| **group\*time** | **2** | **0** | 0 | . | . | . | . |
| **group\*time** | **2** | **2** | 0 | . | . | . | . |
| **group\*time** | **2** | **3** | 0 | . | . | . | . |
| **group\*time** | **2** | **4** | 0 | . | . | . | . |
| **group\*time** | **2** | **6** | 0 | . | . | . | . |

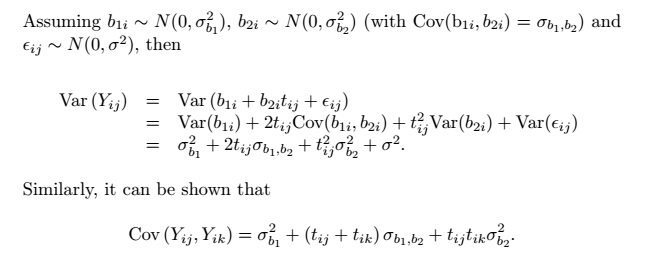
The SE’s for the time effect have decreased, yielding larger Wald tests statistics and smaller p-values. This has also happened for the group-time interaction. For the group effect, the SE’s have increased. This change in SE’s makes me think that the exponential covariance pattern is not quite a good choice.

1. Derive Variance and Covariance in S343 and 347.

Slide 343



Slide 347



[Remark: Mathematical formulas do not need to be 100% correct. But we will check if key ideas are correct.]