



# Answers Key Exercises blavaan

**Note:** Your results should be approximately the same as in this answer sheet. However, they don't have to be exactly the same, because the results can vary because of the sampling from the posterior distribution in the Bayesian analysis.

## Exercise 1: the WAMBS-checklist

### Step 1: Do you understand the priors?

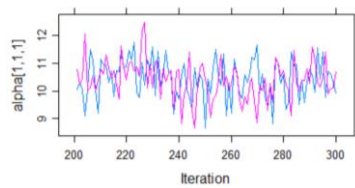
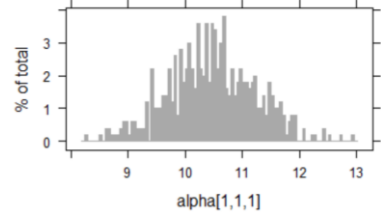
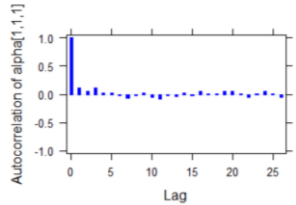
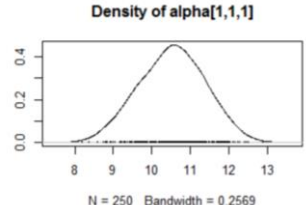
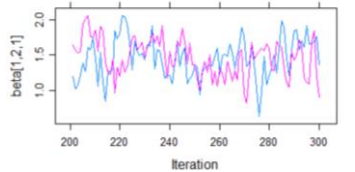
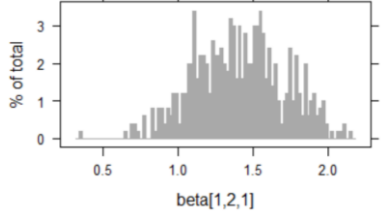
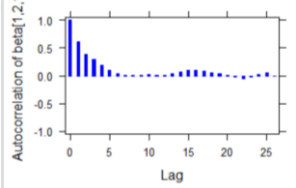
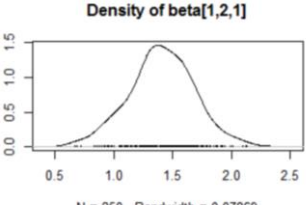
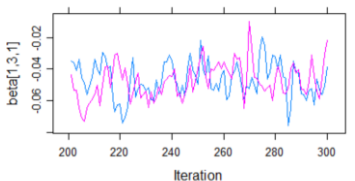
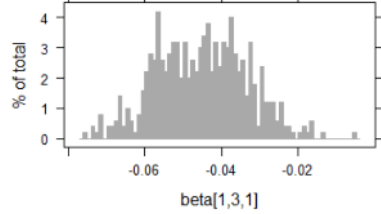
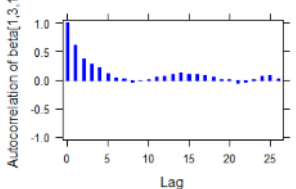
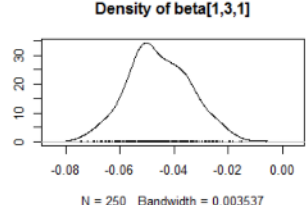
**Table 1.** Do you understand the priors?

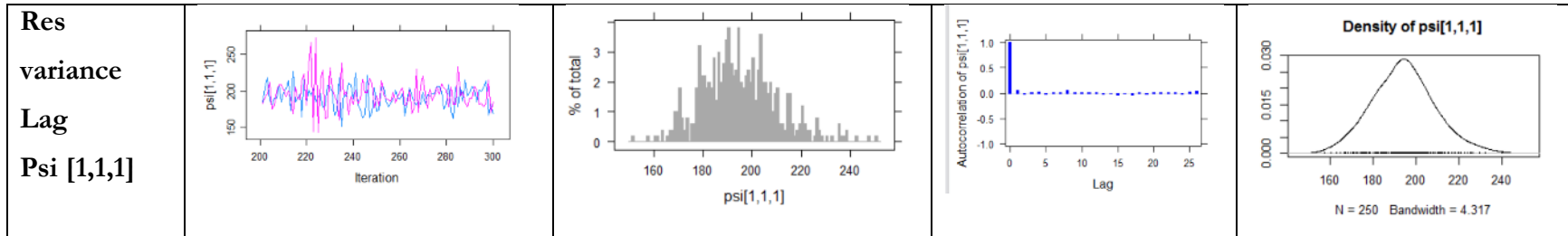
Parameters	Distributional form of the priors (e.g., normal, inverse gamma, etc)	Type of prior (non-, weakly, highly informative)	Source of background information	Picture of Plot	Hyperparameters
Intercept Lag	Normal	Informative	Previous articles		N(10, 5)

Lag on Age	Normal	Weakly informative	<b>Previous articles</b>		$N(.8,5)$
Lag on Age_sq	Normal	Weakly informative	<b>Previous articles</b>		$N(0,10)$
Res variance Lag	Inverse Gamma	Informative	<b>Previous articles</b>		$IG(.5,.5)$

**Step 2: Run the model and check for convergence**

Table 2.

Parameters	Trace plot	Histogram	Autocorrelation Plots	Kernel density plot*
Intercept Lag Alpha[1,1,1]				
Lag on age Beta[1,2,1]				
Lag on age_sq Beta[1,3,1]				



\*The second code of step 6 is used for the Kernel density plot. The first option is also possible, but shows two separate chains in one plot. The second option summarizes these chains in the plot.

### **Step 3: Doubling the number of iterations and convergence diagnostics**

#### **Did the model reach convergence?**

- The Gelman-Rubin Diagnostic shows the psrf values, which all should be close to 1. If they aren't close to 1, you should use more iterations.
- The Geweke Diagnostic shows the z-scores, which should be  $<1.96$ . If this is not the case, you should increase the number of iterations. In the plots you should check how often values exceed the boundary lines of the z-scores. Scores above 1.96 mean that the two portions of the chain significantly differ and full chain convergence was not obtained.
- Your conclusion depends on the diagnostics: if the results are not satisfactory yet, you can conclude that the model didn't converge and therefore you should increase the number of iterations.

After doubling of the number of iterations:

- Check the Gelman-Rubin and Geweke Diagnostics again and compare them with the previous step. Did the situation improve? If not, increase the number of iterations again.
- The bias shouldn't be above  $\pm 10\%$ .

#### **Step 4: Does the posterior distribution histogram have enough precision?**

- The histograms look smooth and have no gaps or other abnormalities. Based on this, adding more iterations is not necessary. However, if you are not satisfied, you can improve the number of iterations again.

#### **Step 5: Do the chains exhibit a strong degree of autocorrelation?**

- There is a strong degree of autocorrelation for the two regression coefficients in the beginning. Based on this, you can conclude that the number of iterations should be increased.

### **Step 6: Does the posterior distribution make substantive sense?**

- The means of the distributions are relatively close to the prior means and the standard deviations are larger than the standard deviations of the posteriors. Only for the regression coefficients the density is not perfectly smooth.

### **Step 7: Do different specifications of the multivariate variance priors influence the results?**

- Looking at the bias percentages, it doesn't seem that making the prior for the residual variance less informative has much effect on the posterior estimates. This indicates that either the data has a stronger influence on the posterior distribution than the prior does, OR that the data and prior are perfectly aligned even if you make the prior less informative.

### **Step 8: Is there a notable effect of the prior when compared with non-informative priors?**

- The difference between the two models is small, but the priors have an influence on the parameter estimates because the values change, but this effect is small. The bias never exceeds +/- 10%. Since the original priors were based on previous research, I would use those, so we can further our knowledge on this topic, instead of starting from scratch (since Bayes with default priors is almost the same as ML).