Advanced Macroeconometrics - Assignment 3

Group 7 - Unterweger L., Lohre F., Langeder E.

2023-05-17

Exercise 1

Question 1.1

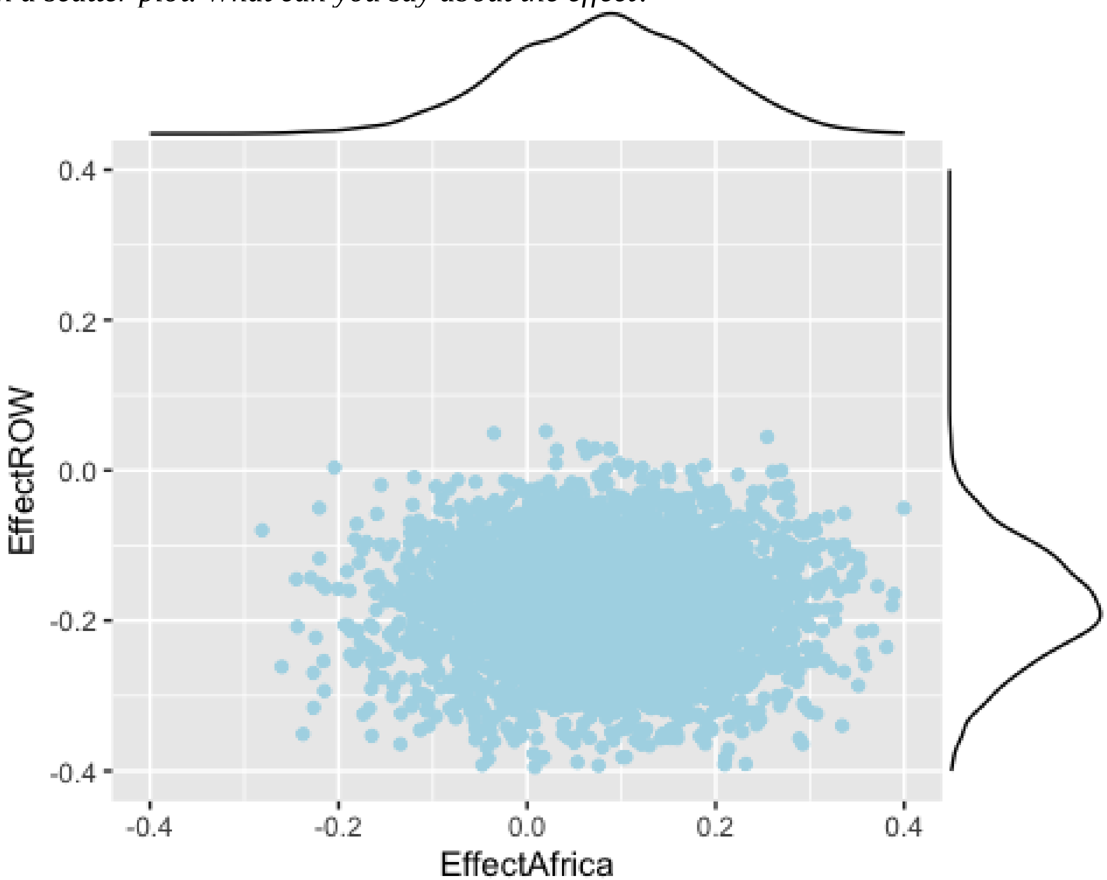
Reproduce the main result, given in Table 1, Column 5 (they use HC1 standard errors), of Nunn and Puga (2012) by fitting a model with the following (interacted) variables:

$$logrgdppc2000 \approx (rugged + dist - coast) \times africa$$

```
##                           Estimate  Est.Error    l-95% CI    u-95% CI
## Intercept                9.3547129 0.13888267  9.08203897  9.62312513
## rugged                  -0.1806744 0.07304451 -0.32478066 -0.04140047
## cont_africa             -1.5055239 0.27854835 -2.03986215 -0.95890505
## dist_coast              -0.6328125 0.17953231 -0.98733391 -0.28435961
## rugged:cont_africa       0.2650344 0.12670866  0.01923812  0.50897612
## cont_africa:dist_coast -0.3729784 0.38491340 -1.13808039  0.35819179
```

Question 1.2

Plot posterior samples of the ruggedness effect in Africa against the effect in the rest of the world in a scatter plot. What can you say about the effect?



From the plot above we can see that the effect in Africa overall tends to be positive, whereas the effect in the rest of the world is largely negative.

Question 1.3

Estimate three additional models — one without the distance to coast, one that uses population in 1400 (use log 1 + pop) instead, and one with both controls.

```
##                       Estimate  Est.Error    l-95% CI   u-95% CI
## Intercept            9.1814316 0.13771792  8.91263808  9.4567547
## rugged              -0.1834696 0.07543635 -0.33268382 -0.0383264
## cont_africa         -1.8370679 0.22251452 -2.25930043 -1.4031075
## rugged:cont_africa   0.3445869 0.13023603  0.09063579  0.5997657

##                             Estimate  Est.Error    l-95% CI     u-95% CI
## Intercept                 9.66434605 0.42909746  8.80210231 10.466585335
## rugged                   -0.19648505 0.07665623 -0.34896846 -0.043433391
## log1Ppop_1400            -0.03575844 0.03337316 -0.10020871  0.030292063
## cont_africa              -0.67397390 0.60516507 -1.86817456  0.520397394
## rugged:cont_africa        0.18443538 0.13565838 -0.08036944  0.459511340
## log1Ppop_1400:cont_africa -0.08881961 0.04480776 -0.17975537 -0.001727031
```

```
##                              Estimate  Est.Error    l-95% CI     u-95% CI
## Intercept                  9.407358848 0.42352341  8.58361278 10.248757519
## rugged                    -0.191325322 0.07515453 -0.33605635 -0.044502104
## log1Ppop_1400             -0.001980979 0.03468996 -0.07032477  0.067194642
## dist_coast                -0.626186138 0.19017919 -0.99725930 -0.248784490
## cont_africa               -0.474492780 0.58634773 -1.60023377  0.690483457
## rugged:cont_africa         0.169997328 0.13481057 -0.09618133  0.437143597
## log1Ppop_1400:cont_africa -0.093633607 0.04737387 -0.18791167 -0.001299089
## dist_coast:cont_africa    -0.018297581 0.40350793 -0.83191630  0.777699719
```

Question 1.4

Discuss (conceptually different) approaches to selecting one of these models for inference. Hint: Consider the difference between causal inference and other inference tasks.

```
##     Model Marginal_Likelihood     WAIC      LOO
## 1 Model 2           -244.5005 469.3148 469.5012
## 2 Model 3           -243.1897 458.7051 458.8293
## 3 Model 4           -239.5196 449.2275 449.4251
```

To select the most suitable model for inference, we can consider several methods. Firstly, comparing the marginal likelihoods of each model can provide insights into their relative fit, where higher values indicate better model fit. Additionally, the Widely Applicable Information Criterion (WAIC) and leave-one-out cross-validation (LOO) can be employed. LOO serves as a measure of the model's predictive accuracy by approximating the leave-one-out predictive density using samples from the posterior distribution obtained through Markov chain Monte Carlo (MCMC) methods. It calculates the log pointwise predictive density for each observation in the dataset, representing the expected log-likelihood of the observation given the rest of the data. Lower values of WAIC and LOO indicate better out-of-sample predictive performance. Thus, by considering these measures, we can assess the models' predictive abilities and make an informed decision.

However, the optimal model selection depends on the analysis objective, where different aspects may hold varying importance. For causal inference, it is crucial to evaluate if the models are designed to address the specific causal question at hand and adhere to appropriate causal assumptions. Validity of estimates should be assessed by examining assumptions, study design, and potential bias sources. Interpretability of estimated causal effects should align with the addressed question. On the other hand, for model fit and prediction, evaluating goodness of fit to observed data and predictive accuracy on training and test data is essential. Model complexity should, thus, be balanced with available data to avoid overfitting. Striking a balance between causal inference and model fit/prediction is crucial, as accurate predictions may not capture causal relationships accurately, and comprehensive causal models may not have the best fit or predictive performance. Ultimately, selecting the best model requires careful consideration of both causal inference and model fit, aligning with specific analysis goals. Assessing models from multiple perspectives and using multiple criteria allows to make an informed decision.

In our example, Bayesian model comparison, using WAIC and LOO, indicates that last estimate model (model 4) is the most suitable. Despite its complexity, for forecasting purposes, a simpler model might however be better suited for causal inference.

Question 1.5

Investigate the sensitivity of the estimates of your model of choice to different prior parameters.

In accordance to the question before, we will proceed with model 4 as our chosen model. To check the sensitivity of our estimates to different prior parameters, we will start by changing the priors for the beta-coefficients.

```
##        PriorAndCoefficient      Model4.1      Model4.2        Model4.3
## 1    Prior for Coefficients Normal (0, 1) Normal (0,10) Normal (20, 10)
## 2                 Intercept         9.427         9.503           9.447
## 3                    rugged        -0.195        -0.197          -0.199
## 4               log1Ppop_1400        -0.003        -0.008          -0.004
## 5                dist_coast        -0.631        -0.645           -0.65
## 6               cont_africa        -0.488        -0.724          -0.616
## 7        rugged:cont_africa         0.172         0.199           0.196
## 8 log1Ppop_1400:cont_africa        -0.093        -0.077          -0.087
## 9    dist_coast:cont_africa        -0.008         0.001            0.04
##         Model4.4
## 1 Normal (-10,4)
## 2          9.623
## 3         -0.202
## 4         -0.017
## 5         -0.635
## 6         -0.993
## 7          0.218
## 8         -0.054
## 9         -0.118
```

We compare model 4 with 4 different normal priors, namely, N(0,1), N(0,10), N(20,10) and N(-10,4). We can observe that the coefficient of interest, rugged:cont_africa, is the lowest for the N(0,1) prior (namely, coef. = 0.172). For the N(-10,4) the coefficient is the highest (0.218). For the other coefficients we can see that the values change slightly with different priors. This implies that they are not very sensitive, even if we use implausible priors.

Now we perform a similar comparison by changing the prior parameters of the inverse gamma distribution.

```
##        PriorAndCoefficient  Model4.5 Model4.6      Model4.7   Model4.8
## 1    Prior for Coefficients IG (2, 1) IG (1,1) IG (0.5, 0.5) IG (20,10)
## 2                 Intercept     9.406      9.4         9.403      9.406
## 3                    rugged    -0.194   -0.195        -0.196     -0.194
## 4             log1Ppop_1400    -0.001   -0.001        -0.001     -0.001
## 5                dist_coast    -0.629   -0.631         -0.63      -0.63
## 6               cont_africa    -0.476   -0.451        -0.473     -0.467
## 7        rugged:cont_africa     0.174    0.171         0.176      0.171
## 8 log1Ppop_1400:cont_africa    -0.094   -0.096        -0.095     -0.095
## 9    dist_coast:cont_africa    -0.014   -0.006         0.006          0
```

We compare model 4 with 4 different inverse gamma priors, namely, IG(2,1), IG(1,1), IG(0.5,0.5) and IG(20,10). We can observe that the coefficient of interest, rugged:cont_africa, is very similar across all IG priors. For the other coefficients we can see that the values hardly change except for the last variable: dist_coast:cont_africa. There we can see that the values change quite a bit for different IG priors.

Question 1.6

Compare the ML using different prior parameters, including prior variances: $\Sigma_0 = v_p I$ *with* $v_p \in$ 0.0001,0.01,1,100.

```
##                  Model marginalLikelihood
## 1 Model - sigma=0.0001          -260.6194
## 2   Model - sigma=0.01          -235.2749
## 3      Model - sigma=1          -239.5307
## 4    Model - sigma=100          -254.8964
```

The marginal likelihood seems sensitive to a change in priors. We can see that the ML increases as the variance decreases.

Question 1.7

Compare the three models using Bayes factors, and explain how they depend on the model prior.

```
##             Model BayesFactor
## 1 Model 2/Model 3 0.268583165
## 2 Model 3/Model 4 0.025442688
## 3 Model 2/Model 4 0.006885775
```

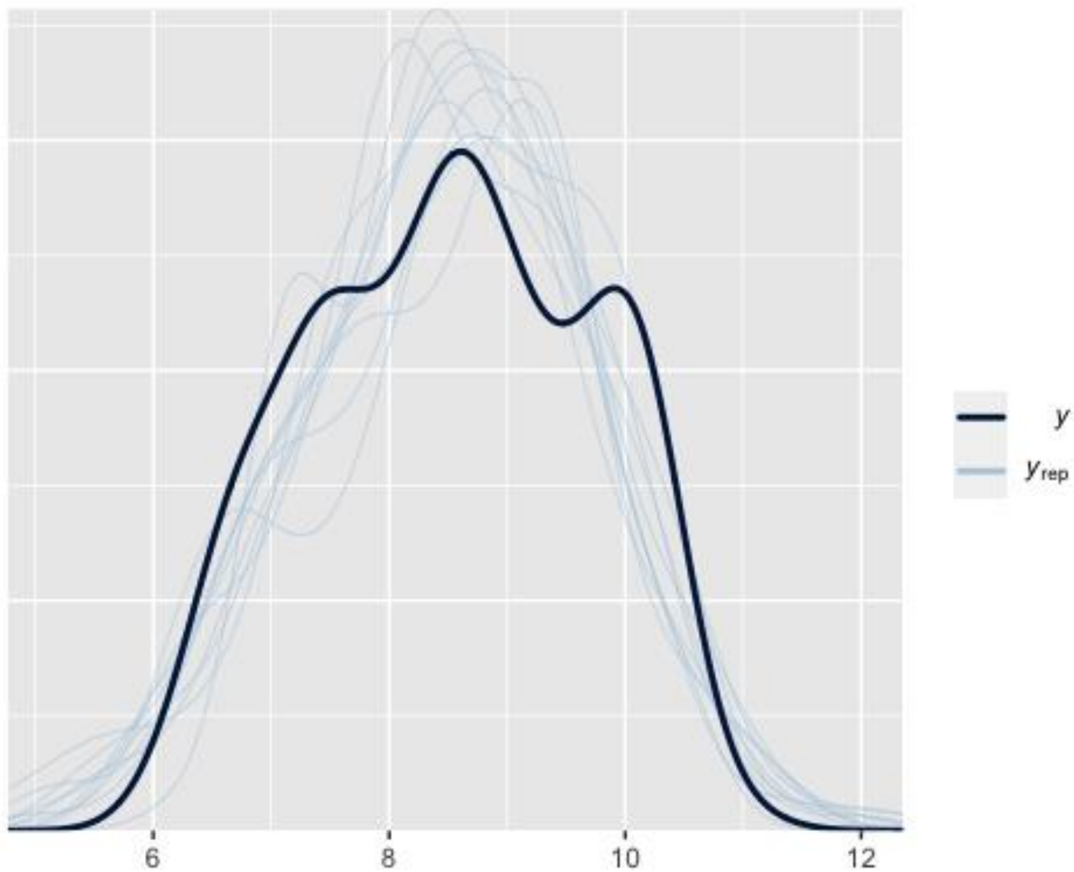
From the table above and in accordance to Andraszewicz et al. (2015) we can see that the Bayes Factor for model 2 vs. model 3 is BF2_3$bf, which means that there is weak but little evidence that supports model 3. The Bayes factor comparing model 3 and model 4 is BF3_4$bf and thus provides substantial evidence towards model 4. The comparison between model 2 and model 4 supports the previous results as it results in a bayes factor BF2_4$bf. INFLUENCE OF PRIORS?

Question 1.8

The posterior predictive density allows us to quantify uncertainty around predictions. To implement this, obtain posterior draws from the model of your choice and use them to simulate predictions. Visualize the predictive uncertainty around a subset of the model.

Like before, we use model 4 our choice.

```
## Using 10 posterior draws for ppc type 'dens_overlay' by default.
```



This density plot compares the observed data to the simulated data to identify potential discrepancies between the model and the data. We can see that there is quite of uncertainty in the simulated data, as the draws are more or less far from the dark line.

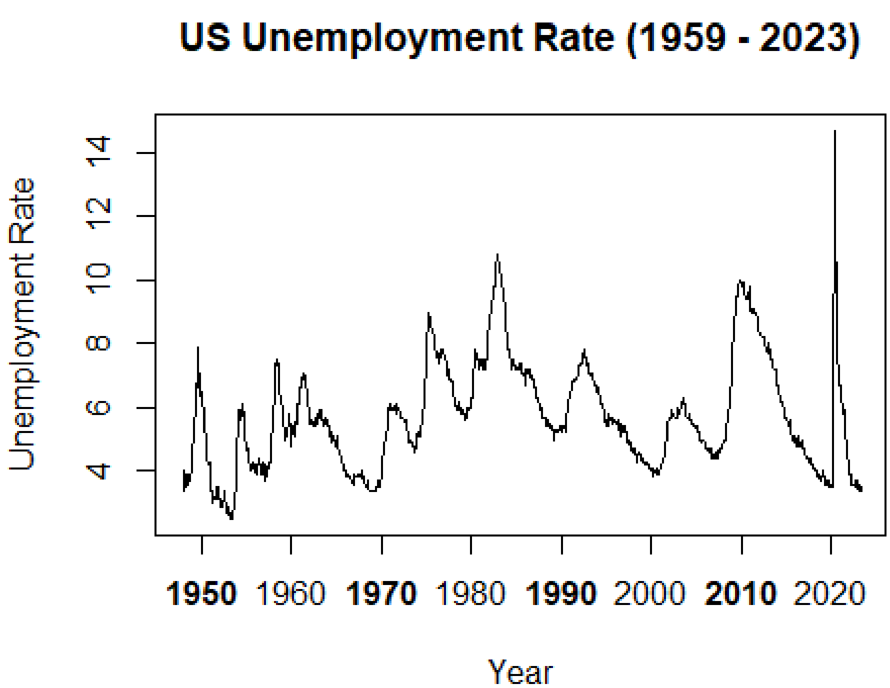
Exercise 2

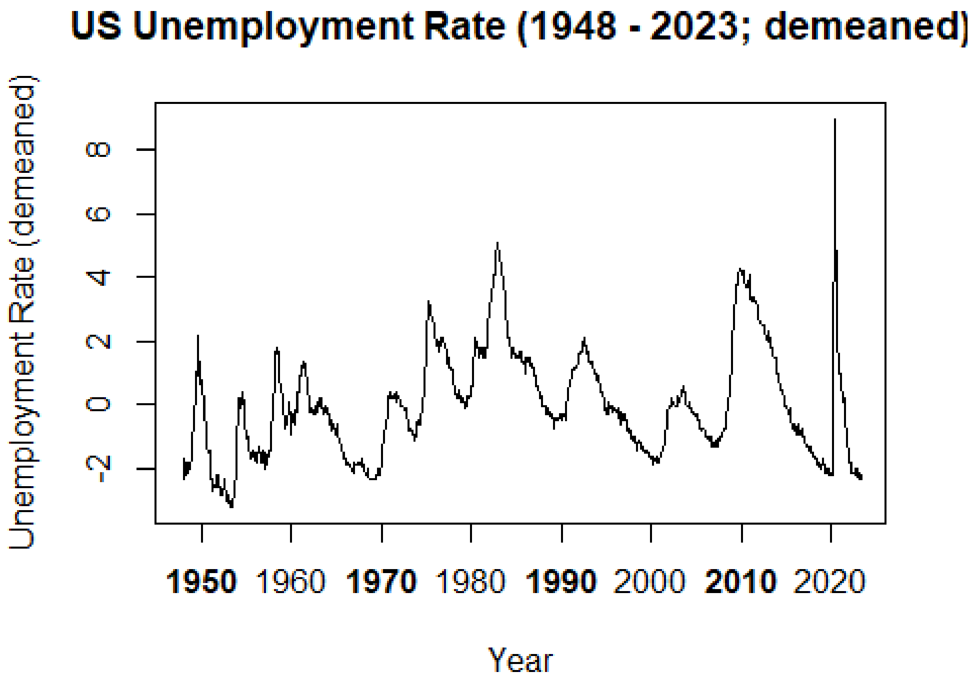
Download the time series on US unemployment rates (mnemonic: UNRATE) from the FRED data base. This is a monthly (seasonally adjusted) time series of unemployment rates running from 1948M1 until today.

```
##         DATE UNRATE
## 1 1948-01-01    3.4
## 2 1948-02-01    3.8
## 3 1948-03-01    4.0
## 4 1948-04-01    3.9
## 5 1948-05-01    3.5
## 6 1948-06-01    3.6
```

2.1 *Demean the time series prior to analysis.*

The original time series as well as the demeaned time series can be found here:





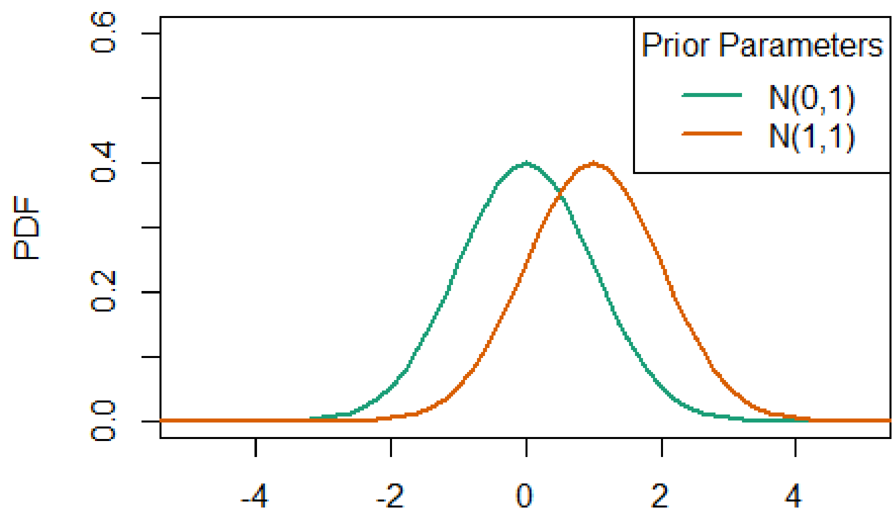
2.2 *Fit an AR(1) process to this time series using the following priors on the autoregressive coefficient:*

Using the stan_sarima() function from the bayesforecast package, we estimate the models.

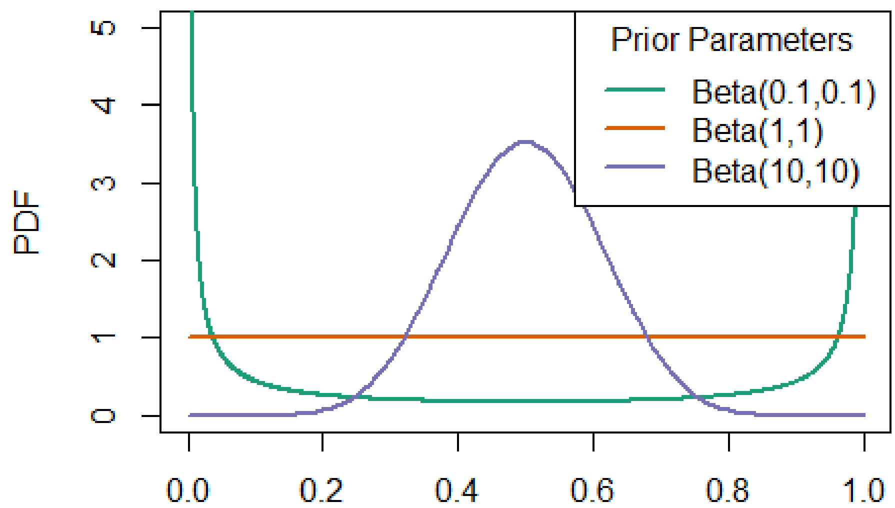
```
AR1_N01 <- stan_sarima(ts = unrate_demean_ts, order = c(1,0,0), prior_ar = ba
yesforecast::normal(0,1))
AR1_N11 <- stan_sarima(ts = unrate_demean_ts, order = c(1,0,0), prior_ar = ba
yesforecast::normal(1,1))
AR1_B0101 <- stan_sarima(ts = unrate_demean_ts, order = c(1,0,0), prior_ar =
bayesforecast::beta(0.1,0.1))
AR1_B11 <- stan_sarima(ts = unrate_demean_ts, order = c(1,0,0), prior_ar = ba
yesforecast::beta(1,1))
AR1_B1010 <- stan_sarima(ts = unrate_demean_ts, order = c(1,0,0), prior_ar =
bayesforecast::beta(10,10))
```

The estimates of the models are given in the appendix.

2.3 Plot the prior distributions and give an intuition for the implied prior assumption for phi.



The intuition for the implied prior assumption would be that the phi's come from a standard normal distribution (mean 0, variance 1) or from a normal distribution with mean 1 and variance 1.

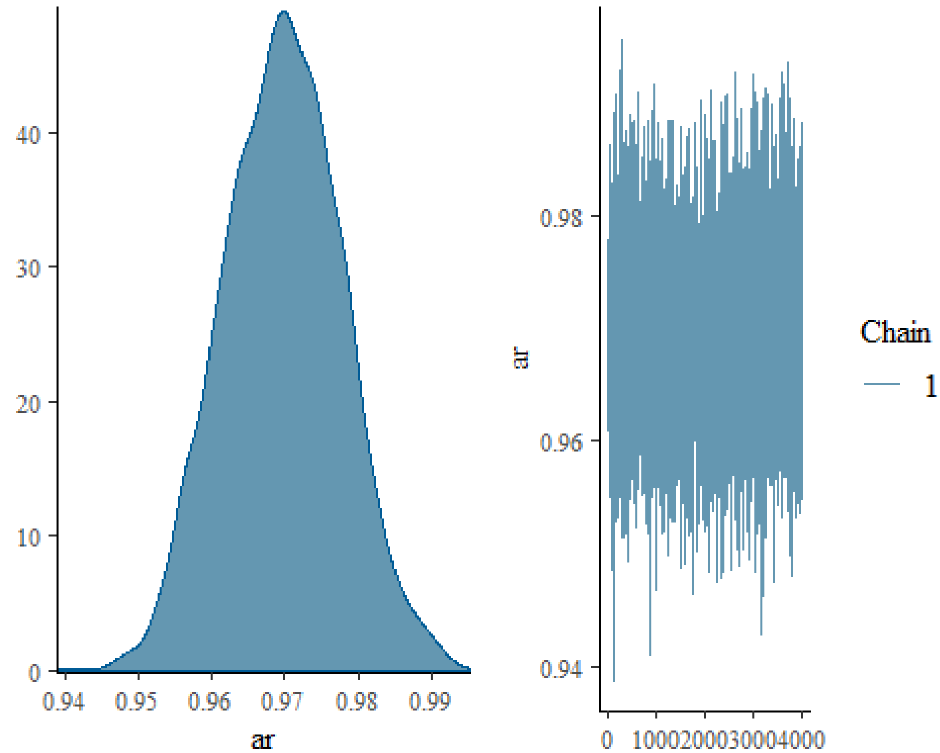


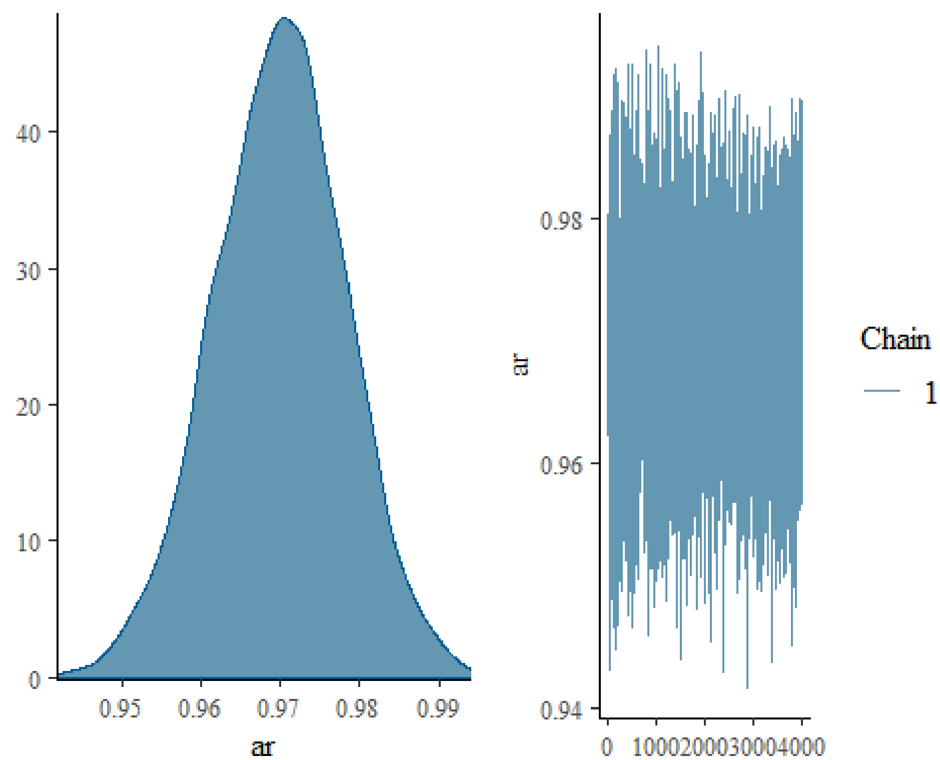
The intuition for the implied prior assumption for the Beta (0.1,0.1) distribution would be that the phi's follow a symmetric distribution where the most weight lies on the tails and decreasing weight moving to the middle.

The intuition for the implied prior assumption for the Beta (1,1) distribution would be that the phi's are basically uniformly distributed from 0 to 1.

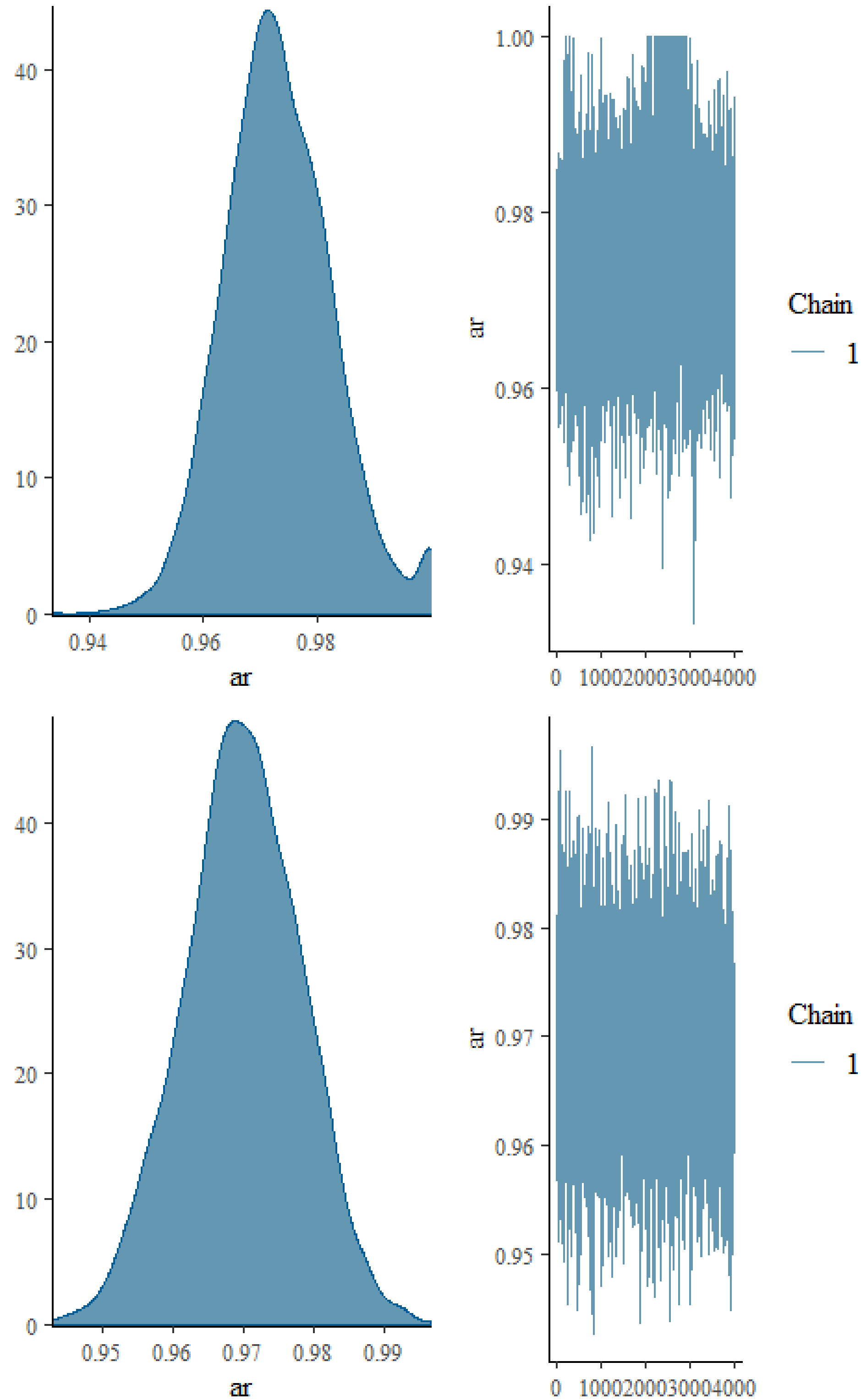
The intuition for the implied prior assumption for the Beta (10,10) distribution would be that the phi's are distributed like a normal distribution with mean 0.5 and a variance of 0.1.

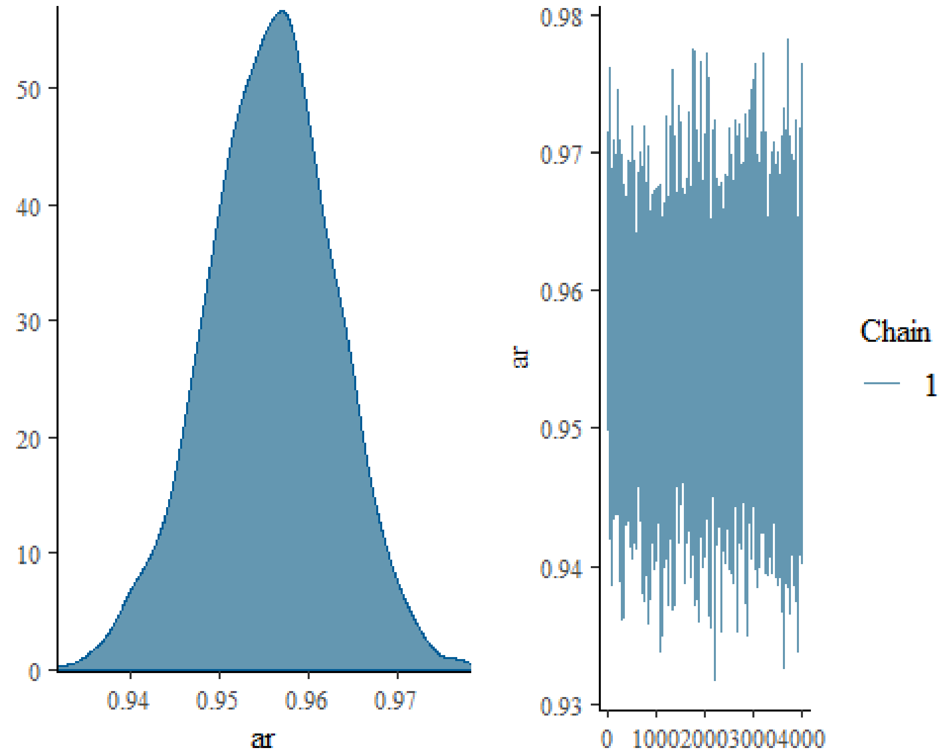
2.4 Plot and interpret the posterior distributions of phi - are they sensitive to the prior?





The posterior distributions for the normal priors look nearly the same. Hence, the difference between those two normal priors is not large and, thus, they are not considered sensitive to the prior.





The posterior distributions of the three beta priors look differently compared to those of the normal priors. The weight close to zero moves away as the parameters of the beta distribution increase. Hence, there is almost no weight around zero for the Beta (10,10) distribution. Nevertheless, the differences do not look substantial and, hence, we also conclude that the posterior distributions of phi are not sensitive to the priors.

Appendix

Esimates of AR(1) models

```
AR1_N01

##
## y ~ Sarima(1,0,0)
## 904 observations and 1 dimension
## Differences: 0 seasonal Differences: 0
## Current observations: 904
##
##              mean     se       5%      95%      ess   Rhat
## mu0      -0.0025 0.0002  -0.0264   0.0208 3811.433 1.0000
## sigma0    0.4227 0.0002   0.4061   0.4397 3782.611 1.0025
## ar        0.9697 0.0001   0.9566   0.9826 4007.139 1.0007
## loglik -503.8692 0.0196 -506.2950 -502.5473 3452.632 1.0024
##
##  Samples were drawn using sampling(NUTS). For each parameter, ess
##  is the effective sample size, and Rhat is the potential
##  scale reduction factor on split chains (at convergence, Rhat = 1).

AR1_N11

##
## y ~ Sarima(1,0,0)
## 904 observations and 1 dimension
## Differences: 0 seasonal Differences: 0
## Current observations: 904
##
##              mean     se       5%      95%      ess   Rhat
## mu0      -0.0025 0.0002  -0.0257   0.0201 4039.481 1.0001
## sigma0    0.4224 0.0002   0.4060   0.4392 3969.520 1.0004
## ar        0.9698 0.0001   0.9561   0.9831 3866.465 1.0001
## loglik -503.8752 0.0190 -506.2608 -502.5462 3681.047 0.9999
##
##  Samples were drawn using sampling(NUTS). For each parameter, ess
##  is the effective sample size, and Rhat is the potential
##  scale reduction factor on split chains (at convergence, Rhat = 1).

AR1_B0101

##
## y ~ Sarima(1,0,0)
## 904 observations and 1 dimension
## Differences: 0 seasonal Differences: 0
## Current observations: 904
##
##              mean     se       5%      95%      ess   Rhat
## mu0      -0.0027 0.0002  -0.0248   0.0207 3907.073 1.0027
## sigma0    0.4229 0.0002   0.4070   0.4396 3922.796 0.9998
## ar        0.9732 0.0002   0.9585   0.9894 3818.148 1.0036
## loglik -504.1052 0.0244 -507.2328 -502.5452 3881.328 1.0061
```

```
##
##  Samples were drawn using sampling(NUTS). For each parameter, ess
##  is the effective sample size, and Rhat is the potential
##  scale reduction factor on split chains (at convergence, Rhat = 1).
```

AR1_B11

```
##
## y ~ Sarima(1,0,0)
## 904 observations and 1 dimension
## Differences: 0 seasonal Differences: 0
## Current observations: 904
##
##              mean     se        5%       95%      ess   Rhat
## mu0      -0.0026 0.0002   -0.0253    0.0199 4151.714 1.0001
## sigma0    0.4225 0.0002    0.4062    0.4395 4052.530 1.0000
## ar        0.9697 0.0001    0.9557    0.9830 4020.812 1.0000
## loglik -503.8720 0.0191 -506.2856 -502.5554 4119.662 0.9999
##
##  Samples were drawn using sampling(NUTS). For each parameter, ess
##  is the effective sample size, and Rhat is the potential
##  scale reduction factor on split chains (at convergence, Rhat = 1).
```

AR1_B1010

```
##
## y ~ Sarima(1,0,0)
## 904 observations and 1 dimension
## Differences: 0 seasonal Differences: 0
## Current observations: 904
##
##              mean     se        5%       95%      ess   Rhat
## mu0      -0.0025 0.0002   -0.0253    0.0208 4128.711 0.9999
## sigma0    0.4235 0.0002    0.4076    0.4402 4115.326 1.0018
## ar        0.9557 0.0001    0.9436    0.9671 4000.271 1.0007
## loglik -505.2039 0.0291 -508.7821 -502.9023 3914.403 0.9998
##
##  Samples were drawn using sampling(NUTS). For each parameter, ess
##  is the effective sample size, and Rhat is the potential
##  scale reduction factor on split chains (at convergence, Rhat = 1).
```